DOCUMENT RESUME

ED 027 514 AL 001 508

By-Labov, William

Contraction, Deletion, and Inherent Variability of the English Copula.

Pub Date 22 Oct 68

Note-69p.; Paper presented at the Annual Meeting of the Linguistic Society of America, Chicago, Illinois, December 1967.

EDRS Price MF-$0.50 HC-$3.55

Descriptors-Cultural Context, Cultural Differences, *English, Generative Grammar, *Language Research, Linguistic Competence, Linguistic Patterns, Linguistic Performance, Linguistic Theory, *Negro Dialects, Nonstandard Dialects, Phonology, *Sociolinguistics, Structural Analysis, Tenl, *Verbs

Identifiers-Copula, *Generative Phonology

This report presents some of the findings of several years research on the relations between the non-standard English used by Negro speakers in various urban ghetto areas (NNE) and standard English (SE). The immediate subject is the status of the copula and auxiliary "be" in NNE. The approach to the problem combines the methods of generative grammar and phonology with techniques for the quantiative analysis of systematic variation. The principal data upon which the discussion is based are drawn from long-term studies of six male adolescent and pre-adolescent peer groups in South Central Harlem and 20 working-class adults out of a stratified random sample of 100 adults from the same area. Two white peer groups from the Inwood section of Manhattan provide controls for comparison with white nonstandard English. The author investigates the form and order of the grammatical and phonological rules controlling the appearance of copula and auxiliary "be" and on the basis of his data concludes that contraction and subsequent deletion of a single consonant is the correct explanation. This model for linguistic research demonstrates that data from the speech community can provide a much-needed, sound, empirical base for decisive solutions to linguistic problems. (MK)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

CONTRACTION, DELETION, AND INHERENT VARIABILITY OF THE ENGLISH COPULA

William Labov,
Columbia University

The following report presents some of the findings of several years research on the relations between the non-standard English used by Negro speakers in various urban ghetto areas [NNE] and standard English [SE].[1] The immediate subject will be the status of the copula and auxiliary be in NNE. The approach to the problem will combine the methods of generative grammar and phonology with techniques for the quantitative analysis of systematic variation. The notion 'rule of grammar' will be enlarged to include the formal treatment of inherent variation as a part of linguistic structure. Furthermore, a model will be presented for the decisive solution of abstract questions of rule form and rule relations, based upon the direct study of linguistic behavior.

The findings and analyses presented here incorporate many contributions of Paul Cohen of Columbia University and Joshua Waletzky of Harvard University, to whom I am deeply indebted.

0. The methods used and the nature of the data.

The study of non-standard Negro English provides a strategic research site for the analysis of English structure in general, for it differs from standard English in many subtle and unexpected ways. However, whenever a subordinate (non-standard) dialect is in contact with a superordinate (standard) dialect, it is not possible to investigate the grammar by eliciting intuitive judgments of grammaticality from native speakers. The data gathered by this method will reflect the superordinate dialect more than the one being studied. There-

AL 001 508

fore it is necessary to study the subordinate dialect by more sophisticated methods, observing the use of this dialect in its normal social setting. The principal data upon which the following discussion is based are drawn from long-term studies of six male adolescent and pre-adolescent peer groups in South Central Harlem,[2] and a sub-sample of twenty working-class adults from the same area, drawn from a stratified random sample of 100 adults. In addition, two white peer groups from the Inwood section of upper Manhattan will provide a base for comparison with white non-standard English [WNS].

Our contact with these groups, and our knowledge of their speech and their social relations, were far more extensive than would be obtained from survey interviews, or from tests in a laboratory or classroom situation. The paradigm for investigating the language of these peer groups may be summarized as follows:

(1) The group was located by the field worker--in most cases, Mr. John Lewis, a participant-observer living in the area.

(2) Several individuals, including the leaders of the group, were interviewed in face-to-face situations.

(3) Our staff met with the group on several outings and trips to various parts of the Metropolitan area. Mr. Lewis maintained daily contact with the group, and made notes on group membership and activities.

(4) In several group sessions, multi-track recordings were made of the group in spontaneous interaction; in these sessions, the dominant factors controlling speech are the same as those which operate in every-day conversation.

(5) All of the remaining individuals were interviewed in face-to-face interaction, and in addition, a large number of isolated individuals in the neighborhood (lames) were interviewed.

We can therefore characterize the language used by our subjects in relation to the speech community, more precisely than with isolated individuals selected by chance or for the convenience of the investigator. This knowledge is an essential prerequisite if we are to write grammars for the speech community, and make inferences about the underlying system from the evidence of language behavior. It is particularly necessary for the present study, since the inherent variation attributed to the basic vernacular here is not to be identified with the fluctuations characteristic of "dialect mixture"; we have indeed studied many marginal members and isolated individuals who show such mixtures, but the data to be given below is based upon the language of members integrated into the peer group, in spontaneous interaction with one another.

The quantitative evidence must of course be recorded under the best possible conditions, and the total output of each individual must be transcribed without ambiguity. Multiple track recordings with individual lavaliere microphones for each individual are necessary to achieve this end. There can, of course, be no question of candid recording in long-term work with a given group. The effects of the recording situation are never absent; they are overridden by more powerful social controls which are exerted by the peer group in excited and rapid interaction.

1. *The status of the copula in NNE*. In this first section, the methods of generative grammar will be used to examine the position of the copula and auxiliary *be* in NNE. It is well known that NNE frequently shows the absence of *be* in a variety of syntactic environments such as those shown in (1-12).

[__NP]
(1) She the first one started us off. [35, S.C., #729][3]
(2) Means he a faggot or sump'm like that. [18, Oscar Bros., #570]

[__PA]
(3) He fast in everything he do. [16, Jets, #560]
(4) I know, but he wild, though. [13, T-Birds, #451]

[__Loc]
(5) You out the game. [10, N.Y.C., #362]
(6) We on tape. [16, Chicago, #471]

[__Neg]
(7) But everybody not black. [15, Jets, #524]
(8) They not caught. [11, T-Birds, #429]

[__Ving]
(9) He just feel like he gettin' cripple up from arthritis. [48, N.C., #232]
(10) Boot always comin' over my house to eat, to ax for food. [10, T-Birds, #451]

[__gn]
(11) He gon' try to get up. [12, T-Birds, #451]
(12) 'Cause we, we gon' sneak under the turnstile. [13, Cobras, #488]

These examples of missing be have led several observers to the conclusion that there is no present copula or auxiliary be.[4] This would seem to be a reasonable inference in view of the fact that a great many languages show no present copula--e.g., Hungarian, or Hebrew. The French Creole of the Caribbean shows the same pattern (13-14), and so does the English Creole of the same area (15-16).

(13) mwẽ ã čwizin. (15) I in the kitchen.[5]
(14) mwẽ esit. (16) I here.

The English Creole of Jamaica shows no copula in some of the environments of (1-12), as for example before predicate adjectives (17) and locatives (18).

(17) im sik bad 'she is very sick'[6]
(18) Jan in a hous 'John is in the house.'

Furthermore, the early grammars used by children 18 to 24 months old show no copula, and there seems to be little basis for constructing one in the underlying phrase structure (19-25).

(19) That a lamb. (22) Kathy in there.
(20) That a bear book. (23) Man in blocks.
(21) It a my book. (24) Tiny balls in there.[7]

The suggestion that NNE shows no copula or auxiliary be is therefore plausible in that this is a very common pattern, particularly in languages which may have had considerable contact with and influence on NNE; in this analysis, NNE would differ from SE in a high level rule of the grammar.[8]

The question raised here should not be identified with the question as to whether the copula appears in the phrase structure of SE or NNE. There are many ways to introduce the copula into the early rules of English grammar, and it is not at all necessary that this be done by a phrase structure rule. The rule given by Chomsky in Aspects of the Theory of Syntax[9] shows a copula in the phrase structure (26).

$$\text{(26)} \quad \text{VP} \rightarrow \left\{ \begin{array}{l} \text{Copula} \quad + \quad \text{Predicate} \\ \text{V} \left\{ \begin{array}{l} \text{(NP)} \quad \text{(PP)} \quad \text{(PP)} \quad \text{(Manner)} \\ \text{S}' \\ \text{Predicate} \end{array} \right\} \end{array} \right\}$$

However, Bach's suggestion appears quite reasonable that the copula should be introduced by an early transformation such as (27) whenever it is followed by a bare predicate, since it is plainly predictable in this environment.[10]

$$\text{(27)} \quad T^{ob}_{cop} \qquad \begin{array}{cccc} \text{X} - & \text{Aux} - & \text{Pred} - & \text{Y} \\ 1 & 2 & 3 & 4 \end{array} \longrightarrow 1 \quad 2+\text{be} \quad 3 \quad 4$$

Another possible approach is that of Rosenbaum in Grammar II; here the auxiliary be is introduced by a segmentalization transformation from features of the following element (28) and the copula could plainly be handled by the same device.[11]

$$\text{(28)} \qquad \begin{array}{ccc} \text{X} - & [+\text{prog}]_{\text{VB}} & - \text{Y} \\ 1 & 2 & 3 \end{array} \longrightarrow 1 \quad \begin{bmatrix} +\text{prog} \\ +\text{COP} \end{bmatrix} +2 \quad 3$$

Whichever method we select for treating the copula, the issue is whether NNE has such high level rules as (26), (27) or (28), or whether NNE differs from SE in not having such a rule. The evidence of the following section supports the former alternative.

2. Environments in which forms of be regularly appear in NNE. Despite the fact that the copula and auxiliary be frequently do not appear in NNE in the variety of environments shown in (1-12), there are a wide variety of other environments in which these forms regularly do appear. The following examples are typical of a large number produced by our grammatical searching of many interviews and group sessions. For most of these environments, the

forms of be appear in the overwhelming majority of cases, and contrary examples are extremely rare: in effect, the appearance of be is a categorical rule, and there are no violations of the rule.

The first examples concern other forms of be besides is and are; these forms are rarely deleted. In the past was and were appear regularly:

(29) I was small; I was sump'm about one years o' baby. [12, Aces, #464]
(30) She was likin' me...she was likin' George too. [18, Oscar Bros., #556]

It can be contended that these are simple past tense markers, with no connection with SE be. Similarly, one might argue that the ain't which regularly appears is merely a negative marker:

(31) It ain't no cat can't get in no coop. [15, Cobras, #490]
(32) My sons, they ain't but so big. [26, N.Y.C., #840]

However, a simple negative not frequently appears as in (7-8), evidently the representative of the negative without the copula. If ain't does not represent is plus not, then we must conclude that there are two negative markers in free variation, or search for some possible semantic difference between They not black and They ain't black.[12]

In the first person, the form I'm is regularly found.

(33) I'm tired, Jeannette. [48, N.C., #232]
(34) I'm not no strong drinker. [15, N.Y.C., #YH44]

This form occurs with overwhelming frequency, despite the fact that it is possible to find rare instances of plain I, I is, or even I'm is. If the task of writing a grammar for a non-standard speech community is that of finding the regular linguistic patterns, we must conclude that the form I'm, which occurs in well over 99°/o of the cases, represents the pattern here.

The case of i's, tha's and wha's provides another case of words in which the copula is frequently represented.

(35) I's a real light yellow color. [15, Cobras, #490]
(36) Tha's my daily routine: women. [14, Cobras, #497]
(37) Wha's a virgin? [12, Jets, #637]

c ..

Accession List checked 4/21/69 PK

I

Accession Number AL 001 508

Date received 11/1/68

No. of copies 1

Sent by author

Solicited? no

Evaluation (Date & Initial)

RIE ✓ (ACW)

CH only

Reject (state why)

* * * * * *

Published?(state in what form):

no

Copyrighted by:

Permission status:

OK

CE sponsored?

Reproduction Level I

Initials/date JT 11/4/68

Abstracted/Indexed:

amm 4/11/68

② MK 2 4/22/69 time units

Snt to NAR on 5/16/69

Subject cards

Descriptor cards

OE doc

Snt. by

author

COMMENTS

author ~~intends~~ "to submit paper to Language in the near future, but there will probably be considered delay".

While we occasionally do get plain *it* as in *It always somebody tougher than you are*, these forms [ɪs], [đæs], and [wʌs] are again found in the great majority of cases, and assume considerable significance for the final analysis of the rule which operates in (1-12).

We also find the form *be* without exception wherever the SE copula would follow a modal or appear in the infinitive form.

(38) You got to be good, Rednall! [15, Jets, #524]
(39) His wife is suppos' a be gettin' money for this child. [48, N.C., #232]

It would seem obvious that the declarative form *You good, Rednall!* corresponds to the modal form (38). There is no way to convert *You good* into **You got to good* without realizing the underlying *be*.[13] The same situation prevails with imperatives.

(40) Be cool, brothers! [15, Jets, #524]
(41) Don't be messin' with my old lady! [16, Jets, #560]

We will now consider environments in which the forms *is* and *are*, which do not appear in (1-12), do appear regularly in NNE. Under emphasis, we find

(42) Allah *is* God. [16, Cobras, #648]
(43) He *is* a expert. [12, T-Birds, #396]

The finite forms of *be* also appear in yes-no questions, as in (44-46):

(44) "Is he dead? is he dead?" "Count the bullet holes in his mother-fucking head." [16, Jets, #560][14]
(45) Are you down? [13, Jets, #497]
(46) Are you gon' give us some pussy? [13, Jets, #632]

We also obtain yes-no questions without *is* and *are*; the problem of the question transformation, and the base forms of questions must be considered elsewhere. But in the large number of cases where *is* and *are* do appear in questions, we must relate them to underlying declarative sentences with copula *be*. The examples chosen here are deliberately selected to show that these are vernacular forms: to explain these examples as "dialect mixture" or as importations from standard English would be an extremely unlikely hypothesis.

In the case of tag questions, as in (47), the finite forms of *be* are required:

(47) Is that a shock? or is it not? [13, Cobras, #493]

Again we find that *is* occurs in quotations from the most excited and spontaneous interaction in group sessions.

The most interesting set of examples, from the syntactic point of view, are those in which we find *is* and *are* in clause-final position, as the result of several transformational processes. In elliptical responses:

(48) (You ain't the best sounder, Eddie!)
I ain't! He is! [12, Cobras, #489]

After ellipsis in comparative constructions:

(49) He is better than the girls is, now. [35, S.C., #729]
(50) It always somebody tougher than you are. [25, Fla., #825]

In embedded questions, after WH-attraction:

(51) That's what he is: a brother. [14, Cobras, #492]
(52) I don't care what you are. [16, Jets, #580]
(53) Do you see where that person is? [15, N.Y.C., YH35]

In all of these frequent forms, we find the finite forms *is* and *are* without exception: the alternatives without the copula or auxiliary simply do not exist.

With sufficient ingenuity, it is possible to provide an *ad hoc* explanation for each one of the cases in this section, and claim that there is no connection between the forms found there and the sentences of (1-12).[15] However, it will be obvious to all familiar with the logic of transformational grammar that the evidence given here points to the existence of an underlying copula and auxiliary *be* which is deleted in the specific environments of (1-12). The question then remains, by what kind of rule are these finite forms of *be* deleted? Is it a transformational rule which deletes the copula, or a separate set of rules which delete *is* and *are*?. Or is it a phonological rule which operates at a lower level in the grammar? We will now proceed to specify the nature of this deletion rule more precisely.[16]

3. *The general nature of the deletion rule, and its relation to contraction*. First, we can observe a number of signs of phonological influence upon the deletion rule. *Is* and *are* are deleted, but *'m* is not: there are phonological processes which operate upon final [z] and [r] in NNE, but not upon final [m]. *Ain't* and *be* are phonologically distinct from *is* and *are* in that they contain tense vowels which are not reduced to schwa or contracted. *Was* and *were* begin with a consonant which is not generally deleted. *I's* [ɪs], *tha's* [ðæs] and *wha's* [wʌs] are plainly the result of some low-level process of assimilation, which transforms them in such a way that they are protected from the deletion rule. It follows that the deletion rule is ordered after the processes which change *it is* to *i's* [ɪs].

But the most important suggestion which proceeds from the examples of section 2 is the relation between contraction and deletion. We find that the following general principle holds without exception: *wherever SE can contract, NNE can delete is and are, and vice-versa*; *wherever SE cannot contract, NNE cannot delete is and are, and vice-versa*. This intimate relationship between contraction and deletion will be illustrated by the examples below.

3.1. *The rule for contraction of the English auxiliary*. To the best of my knowledge, the rules for SE contraction have never been explored in print in any detail. It will therefore be necessary to look into the conditions under which contraction can occur, and specify the form of the contraction rule, in order to understand its relation to deletion and the form and position of the deletion rule itself.

Just as SE cannot contract in final position, so NNE cannot delete. Examples (54-57) illustrate the parallel:

	SE	NNE
(54)	*He's as nice as he says he's.	*He's as nice as he says he.
(55)	*How beautiful you're!	*How beautiful you!
(56)	Are you going? *I'm.	Are you going? *I.
(57)	*Here I'm.	*Here I.

The patterns shown in the actual data are so absolute that I feel justified in placing asterisks in the NNE column to indicate that the form is impossible, even without asking for intuitive judgments of native speakers. From these examples, it would appear that the rule is simply that contraction is impossible in final position. But (58-61) show that there is more to the matter than this.

(58)	*Who's it?	*Who it?
(59)	Who's IT?	Who IT?
(60)	*What's it?	*What it?
(61)	What's it for?	What it for? Wha's it for?

We cannot say (58) with dummy *it*, although we can say (59) with lexical IT ('the person who is IT in a game'). We cannot say (60), with dummy *it*, but we can say (61), when stressed *for* follows. It would seem then that a stressed syllable must follow the *is* or *are* if it is to be contracted or deleted. Still, (62-64) show that the situation is more complex.

(62)	*He's now.	*He now.
(63)	*He's unfortunately.	*He unfortunately.
(64)	He's unfortunately here.	He unfortunately here.

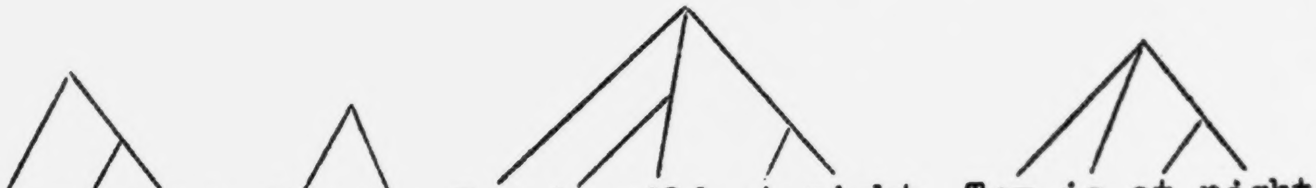
In both (62) and (63), there are stressed forms following the copula, yet we cannot delete or contract. In (64), after the addition of *here*, we can contract and delete. It is evident at this point that the grammatical relations between *is* and *are* and the following elements are important to the rule. Such grammatical relations figure in the stress assignment rules provided by Chomsky and Halle in *Sound Patterns of English*, and these allow us to state the initial conditions which govern contraction.[17] The following set of three rules operate to provide these conditions.

(65) $[V] \rightarrow [\overset{\prime}{V}]/[\overset{\prime}{V}]\ldots[\overset{\prime}{\_}]]_{a}$ Nuclear stress rule

(66) $V \rightarrow [-\text{stress}]/[\_\_, +W, 3\text{stress}]$ Weak word rule

(67) $V \rightarrow ə\ /[\_\_, -\text{tense}, -\text{stress}]$ Vowel reduction

The nuclear stress rule is a cyclical rule which re-assigns primary stress to the last lexical item within each phrase marker, by convention reducing the stress assignment of all other items by one unit. The phrase marker boundaries are then erased, and the rule applies to the next larger phrase. The weak word rule,[17a] provided by me, operates so that weak words--words which can occur with schwa as their only vowel--are reduced to [-stress] from [3 stress], whereas other syllables will be reduced to [-stress] only from [4 stress] or [5 stress], and weaker. The vowel reduction rule (67) is the last rule in the Chomsky-Halle series. Contraction then follows: it is the removal of a schwa which occurs initially in a word before a lone consonant. In the examples given below, the operation of these rules is illustrated.



Rule	Tom	is	wild	Tom	is	Tom	is	wild	at	night	Tom	is	at	night
	1	1	1	1	1	1	1	1		1	1	1		1
(65)		2	1	2	1		2	1			2	2		1
(65)	2	3	1			2	3	2		1				
(66)		-					-							
(67)		əz					əz							
Cont'n		z					z							

Tom's wild.

Tom's wild at night.

In Tom is wild, the nuclear stress rule operates twice, reducing is to [3 stress]; then the weak word rule makes this [-stress], vowel reduction applies, and contraction, yielding Tom's wild. In the elliptical form Tom is, we have only one cycle with full stress on is (or if emphatic stress is placed on Tom, with [2 stress] on is). No contraction is possible. In Tom is wild at night[18] there are again two cycles, and the rules yield Tom's wild at night. But after ellipsis of wild, as in Bill is wild during *the day, and Tom is at night*, the copula *is* is not in construction with *at night*, and there is only one cycle for the nuclear stress rule. Therefore contraction does not apply.

The form of the contraction rule, therefore, will show that it represents the removal of an initial schwa before a lone consonant as in *am*, *is*, *are*; *have*, *has* and *had* will be included after a very general rule removes the initial *h*-; *will* is included, apparently with a lexical alternate without the initial *w*-, since there is no general rule to delete this consonant. But unstressed *as* cannot be contracted, even though it has the requisite phonological form [əz]. We know this because voicing assimilation, which occurs automatically after contraction, does not apply to *as* in *like as not* or *hot as can be*: no matter how ephemeral the schwa seems to be, we do not say [laɪksnɑt] or [hɑtskənbi]. Nor are *his*, *him* or *her* contracted, although the rule which removes the initial *h*- applies to them as well as to *has*, *had*, *have*.

It appears from these examples that contractability may be a lexical property of these verbs or auxiliaries: some variation may be noted in the verb *have*, which is contracted in British English, as in *They'd a great deal of money*, but not in American English. Despite this idiosyncrasy of *have*, we can find a general feature of the context which determines contractability, and shows why *as*, *him*, *his*, *her* do not contract, while both auxiliaries and copula generally do. Contraction requires the presence of the type or tense marker. The critical case is found in *They may have*. This can be written as *They may've*, but the apostrophe only indicates the deletion of the *h*-. Contraction has not applied, as we can tell from the fact that *They may've* does not rhyme with *knave*. When contraction does operate to remove the schwa, we obtain a single syllable: *They've* does rhyme with *knave*. Thus contraction occurs only when the tense or type marker[19] is incorporated in the verb or auxiliary, and the form of the contraction rule has the general shape of (68).

(68) $ə \rightarrow (\emptyset) / \ldots \#\# \_\_ \; C_0^1 \; \#\# \ldots$

The dots imply that there are further constraints upon contraction which will be discussed below. We have developed the contraction rule as far as we can within the framework of categorical, invariant rules. There are further problems, and further constraints upon contraction which can only be handled with an enlarged conception of 'rule of grammar'.

3.2. Relations of order between contraction and deletion. One such further problem concerns the relations between the contraction rule, as generally sketched above, and the deletion rule of NNE. There are four possible relatinns of order between contraction and deletion:

Case 1	Case 2	Case 3	Case 4
1. C 2. D	1. D 2. C	1. $\left\{\begin{matrix}C\\D\end{matrix}\right\}$	1. C(D)
əz→z/.. z→∅/..	əz→∅/.. əz→z/..	əz→$\left\{\begin{matrix}z\\\emptyset\end{matrix}\right\}$/..	əz→ z→∅/..

Case 1 is that contraction occurs first, deletion second. Case 2 is the reverse: deletion first, contraction second. It is apparent from the forms suggested that no particular relation between the two rules is implied by this order; for many reasons, Case 2 will appear the least likely. Case 3 shows deletion and contraction as simultaneous alternates of the same rule, with only one set of environmental constraints. Case 4 has deletion as an extension of contraction--contraction gone wild, as it were--again with only one set of environmental conditions. Our task is now to discriminate among these four possibilities of order, and to specify in detail the form of the deletion rule.

4. Inherent variability of deletion. So far, we have presented the forms (1-12) of section 1 as if this were the pattern of NNE. And indeed, this is the pattern which is most frequently noticed, for it is marked by its deviation from SE. However, deletion of the copula is an inherent variable for all

of the NNE speakers whom we have studied. We will now explore the internal structure of this variable characteristic in order to solve the problems of ordering raised in the preceding section.

The study of variation is necessarily quantitative, and quantitative analysis necessarily involves counting. At first glance, counting would seem to be a simple operation, but even the simplest type of counting raises a number of subtle and difficult problems. The final decision as to what to count is actually the final solution to the problem in hand. This decision is approached only through a long series of exploratory maneuvers.

First, one must identify the total population of utterances in which the feature varies. There are always some parallel cases where the variable feature is not variable at all--as, for example, the environments of (48-53) above where we find that _is_ is never deleted. If all of the environments of (29-53) were included in a quantitative study of the variable deletion rule, the frequency of application of the rule would appear much lower than it actually is; a number of important constraints on variability would be obscured since they would appear to apply to only a small portion of the cases; and the important distinctions between variable and categorical behavior would be lost.

Second, one must decide on the number of variants which can be reliably identified, and set aside those environments in which the distinctions are neutralized for phonetic reasons. In the case of _is_, we decided to isolate full, contracted, and deleted forms, but not to attempt to distinguish the degree of stress or reduction of the vowel in the full form. Furthermore, sentences such as _Boot is seventeen_ must be set aside, since the contracted form cannot be distinguished from the deleted form in [butsɛvntin] or [but·sɛvntin].

Third, one must identify all of the sub-categories which would reasonably be relevant in determining the frequency with which the rule in question applies. In this case, there are many grammatical and phonological characteristics of the preceding and following element which determine the frequency of contraction and deletion of is: few of these can be predicted from any current theory or knowledge about contraction. Such sub-categories emerge from the ongoing analysis, as a result of various fears, suspicions, inspections and analogies. Although there is of course no simple procedure for the isolation of relevant sub-categories, the end result is a set of regular and powerful constraints which operate upon every group and almost every individual. When the three operations outlined above are carried out with any degree of accuracy and linguistic insight, the regularities are so evident that statistical analysis is superfluous.

In this section we will focus upon the quantitative analysis of the forms of is in the environments of (1-12). There are no NNE speakers in our sample, or in our exploratory work in Washington, Philadelphia, Cleveland, Detroit, Chicago or Los Angeles, at any age level, in the most excited and spontaneous interaction, who always (or never) delete is in these environments. Full, contracted and deleted forms are all characteristic of NNE. The contracted (undeleted) form is least typical of NNE, and most characteristic of WNS and SE. On the analogy of the SE and WNS feeling that contracted forms are "natural" and that full forms are "careful", one might be tempted to argue that the full forms are importations from SE in "careful" style. However, as we move from single,-face-to-face interviews to spontaneous group sessions, we find that the percentage of full forms generally increases. The feature which is correlated with style shift from single to group sessions is the ratio of deleted to originally contracted forms--that is, D/D+C. In other words, NNE speakers do not necessarily contract more in excited interaction, but they delete more of the forms which have been contracted. However, these stylistic shifts are minor effects among the pre-adolescent and adolescent peer groups, and only begin to assume importance with the older adolescents and adults.[20]

The single most important constraint on deletion in NNE and upon contraction in SE and NNE is one which we did not expect--whether or not the subject is a pronoun or some other noun phrase. Table 1 and Figure 1 show the percentages of full forms [F], contracted forms [C], and deleted forms [D] for six groups that have been studied closely: the pre-adolescent Thunderbirds, the adolescent Cobras, Jets, and (somewhat older) Oscar Brothers; a sample of one quarter of the working-class adults in the Cobra and Jet areas from the larger random sample of 100 adults; and the combined records of two white working-class groups--pre-adolescent and adolescent--from the Inwood neighborhood of upper Manhattan.

On the left of each figure in Figure 1 is the percentage of full, contracted and deleted forms after noun phrases: on the right, after pronouns. In every case, the percentages of deleted and contracted forms are greater when a pronoun precedes. The upper line of figures shows the pattern for single interviews; the bottom, for group interaction.[21] Though there is a general increase in the ratio of deletion to contraction, the basic pattern is the same in both styles, for all groups.

In these diagrams, deletion is shown as occurring after contraction (Case 1); that is, the total percentage of contracted forms includes those forms which were afterwards deleted. The pattern for contraction shown here is similar for the NNE groups and for the WNS Inwood groups, who do not delete. Contraction and deletion thus respond to the same syntactic constraint. The fact that this pattern repeats regularly in six different groups, in each style, indicates how pervasive and regular such variable constraints are. We are not dealing here with effects which are so erratic or marginal that statistical tests are required to determine whether or not they might have been produced by chance.

The relationship between contraction and deletion can be explored more deeply by considering the effect of the following grammatical category. Again, we find that both rules respond

Figure 1. Per cent of full, contracted and deleted forms of is with pronoun subject vs. other noun phrase subject for six Harlem groups in single and group (casual) style

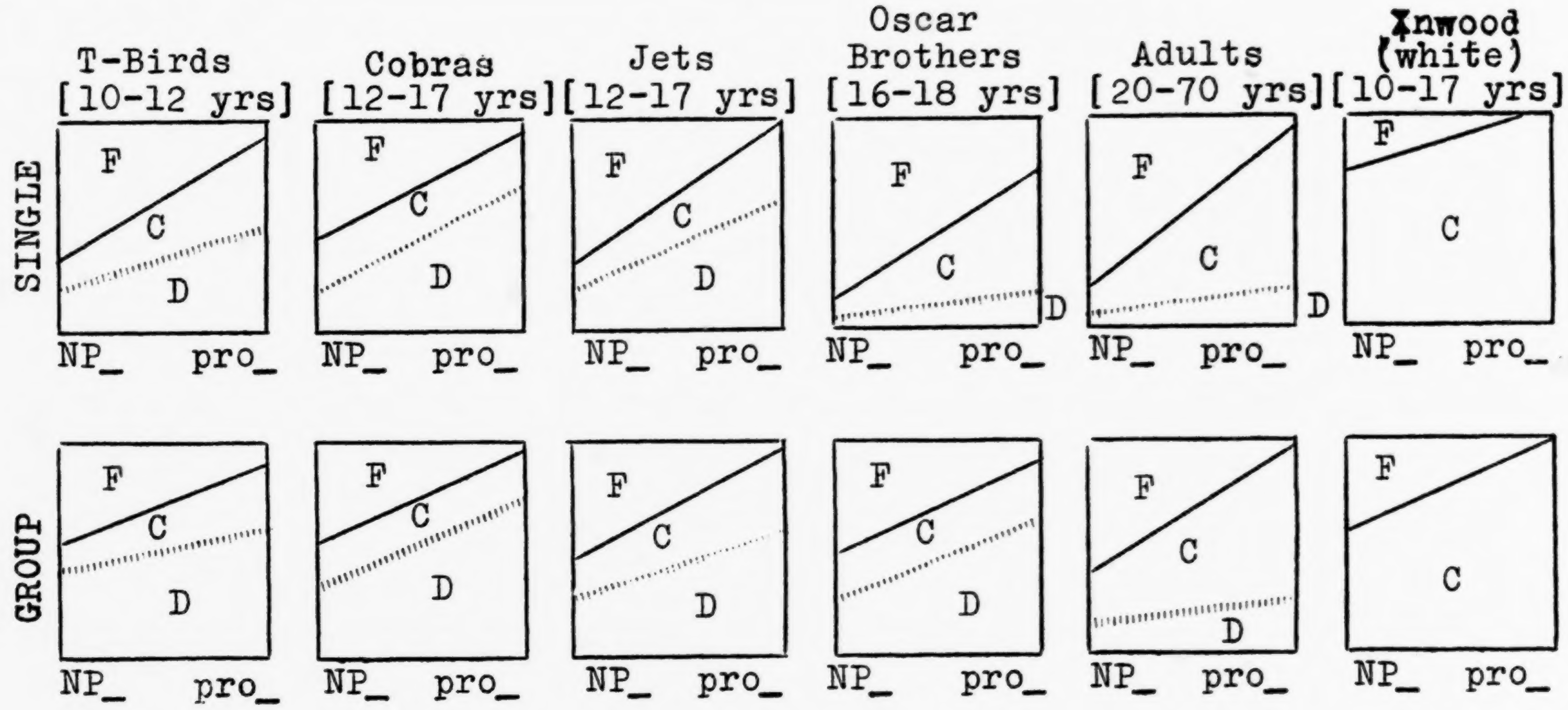
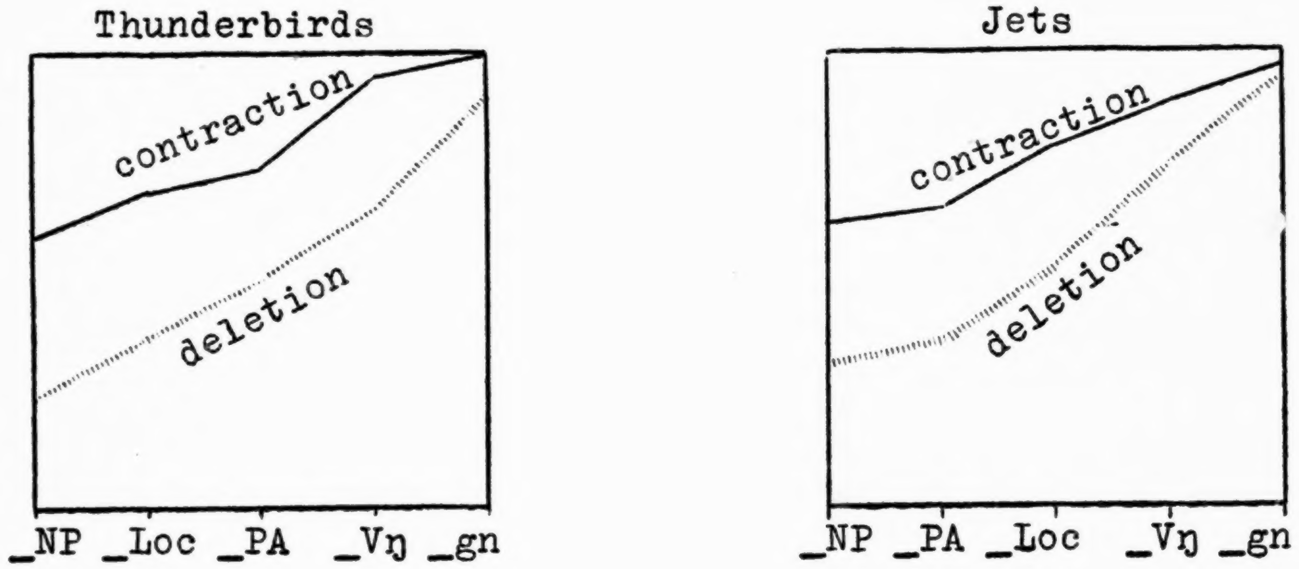


TABLE 1

PER CENT OF FULL, CONTRACTED AND DELETED FORMS OF IS WITH PRONOUN SUBJECT VS. OTHER NOUN PHRASE SUBJECT

	T-Birds		Cobras		Jets		Oscar Bros.		Adults		Inwood	
	NP_	pro_	NP_	pro_	NP_	pro_	NP_	pro_	NP_	pro_	NP_	pro_
SINGLE STYLE												
Full	63	05	56	04	67	00	85	25	75	04	26	00
Cntrd	25	44	26	29	15	39	11	60	17	80	74	100
Dltd	12	51	18	67	18	61	04	15	08	16		
	100	100	100	100	100	100	100	100	100	100	100	100
N:												
Forms	124	212	35	106	145	189	45	47	187	118	54	61
Subjs.	13		9		15		3		17		8	
GROUP STYLE												
Full	44	07	45	00	54	00	51	04	61	01	41	01
Cntrd	15	33	19	23	19	42	23	33	26	72	59	99
Dltd	42	60	36	77	27	58	26	64	14	27		
	101	100	100	100	100	100	100	101	101	100	100	100
N:												
Forms	53	43	85	30	113	75	73	80	170	112	110	81
Subjs.	5		9		11		4		15		7	

Figure 2. Per cent of full, contracted and deleted forms of is according to grammatical category of complement


Thunderbirds

Jets

TABLE 2

PER CENT OF FULL, CONTRACTED AND DELETED FORMS OF IS ACCORDING TO GRAMMATICAL CATEGORY OF COMPLEMENT FOR TWO GROUPS IN ALL NNE STYLES

	_NP	_PA	_Loc	_V+ing	_gonna	
THUNDERBIRDS						(13 subjs.)
Full	40	25	30	04	00	
Contracted	37	27	34	30	12	
Deleted	23	48	36	66	88	
	100	100	100	100	100	
No. forms	210	67	50	46	40	
JETS						(29 subjs.)
Full	37	34	21	07	03	
Contracted	31	30	27	19	03	
Deleted	32	36	52	74	93	
	100	100	100	100	99	
No. forms	373	209	70	91	58	

to the same set of syntactic constraints. Table 2 and Figure 2 show this pattern for the Thunderbirds and the Jets, for single and group styles combined. The relationships shown here[22] are essentially the same for the other groups. The least deletion and contraction take place before a following noun phrase; more occurs before predicate adjectives and locatives; both rules apply with even greater frequency before a following verb with the progressive -ing; and with the highest frequency before the future form gonna. Here contraction is again shown as taking place on the full population of full forms, but the population upon which the deletion rule operates is limited to the pool of forms already contracted.

Figure 2' below shows the consequences of treating contraction and deletion as independent processes. Here the percentage of contraction for the Jets is shown in terms of the actual numbers of contracted forms recorded: the result is a minor tendency which responds in just the opposite way to the syntactic constraints. Furthermore, there is no connection at all between contraction in NNE and contraction in WNS: Figure 3 on the right shows the contraction pattern of the Inwood group, quite similar to the "cumulative" contraction pattern of Figure 2 (indicated on Figure 2' with a dotted line). If, then, we should insist on regarding contraction and deletion as completely unrelated, we would find that the syntactic constraints which operate upon them

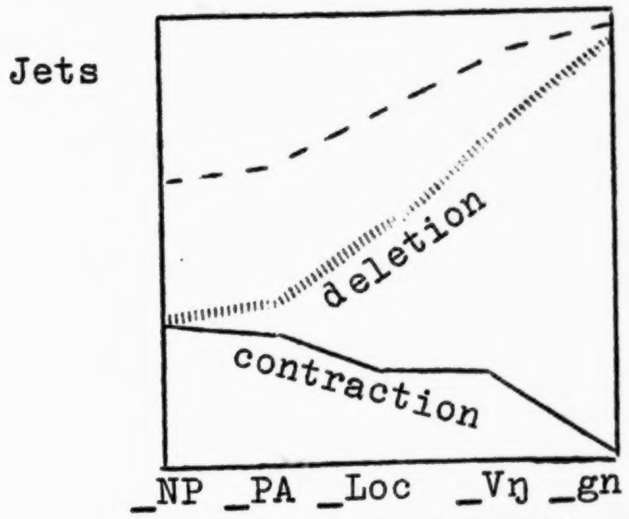


Fig. 2'. Contraction and deletion independent

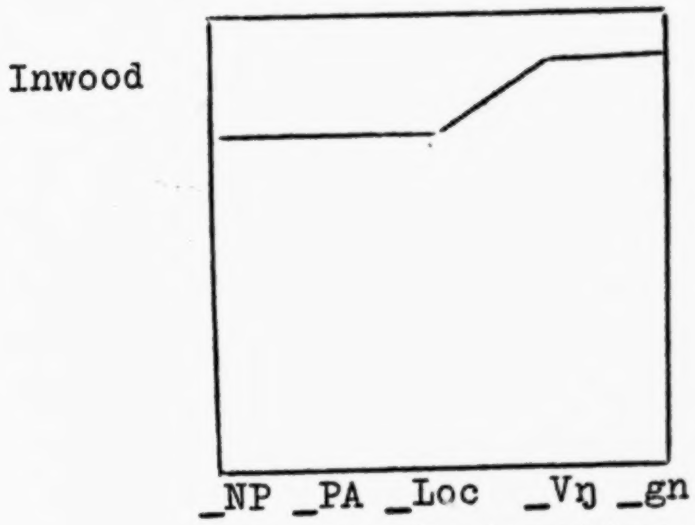


Fig. 3. Contraction for the Inwood groups

have very different effects, and that contraction for NNE has nothing to do with contraction for WNS. This is a very implausible result, and we can proceed upon the assumption that the cumulative diagram of Figure 2 represents the actual situation.

Given these quantitative relations, we can now return to the problem of the particular form of ordering which holds between the contraction and deletion rules. The four cases of possible ordering presented above can now be simplified. Case 2, with deletion first and contraction second, would not fit any of the quantitative results shown above, for there is no reason for the contraction of some undeleted [əz] to be dependent upon the deletion of some other [əz]: that is, it would be quite unreasonable to insist that contraction operates upon a pool of already deleted forms. The other three cases can be represented by the abstract quantitative models of Figures 4a-c below.

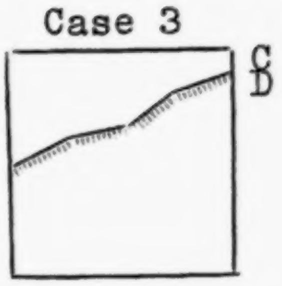


Fig. 4-a

$\text{əz} \rightarrow \overset{z}{\emptyset} /..$

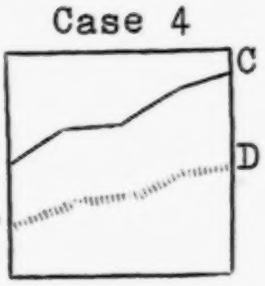


Fig. 4-b

$\text{əz} \rightarrow z \rightarrow \emptyset /..$

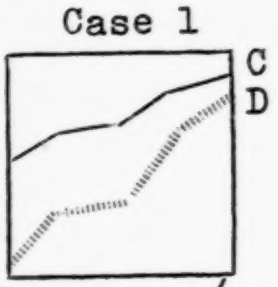


Fig. 4-c

1. $\text{əz} \rightarrow z/..$
2. $z \rightarrow \emptyset/..$

The application of the variable contraction and deletion rules is logically governed by two factors: first, an input variable which sets the overall frequency with which the rule is selected. Secondly, there are variable constraints in the immediate environment which differentiate the frequencies with which the rule applies according to various syntactic and phonological features of the sentence.[23,24] Figures 4a-c represent the quantitative results of various combinations of these factors. For Case 3, with contraction and deletion as alternative right hand members of a single rule, we have

$$\text{əz} \rightarrow \begin{Bmatrix} z \\ \emptyset \end{Bmatrix} / \cdot\ \cdot$$

In this expression, the rule is selected only once, and there is therefore only one variable input and one set of variable constraints. The spectrum of frequencies with which the contraction and deletion rules apply should therefore be the same, as shown in Figure 4-a. If, on the other hand, ddeletion is thought of as an extension of contraction, as in Case 4

$$\text{ə}z \rightarrow z \rightarrow \emptyset \; / \; . \; .$$

we might have two selections and two variable inputs, but only one set of variable constraints. Thus deletion would be a fixed percentage of contraction in all environments--say 50^{o}/o, as suggested by Figure 4-b. The third possibility is that we have two selections (with variable inputs), and two sets of variable constraints. This is in effect equivalent to Case 1, with the rule for contraction applying first and the rule for deletion applying second. Here the quantitative pattern would be that of Figure 4-c, where the variable constraints apply twice. This pattern shows more extreme or exaggerated constraints upon deletion than upon contraction; it is in fact the actual pattern which appears in the empirical data of Figure 2 for both the Thunderbirds and Jets, and one which is repeated for the other peer groups as well[25] We can therefore conclude from this quantitative evidence that contraction and deletion are separate, though similar, rules which apply in that order.

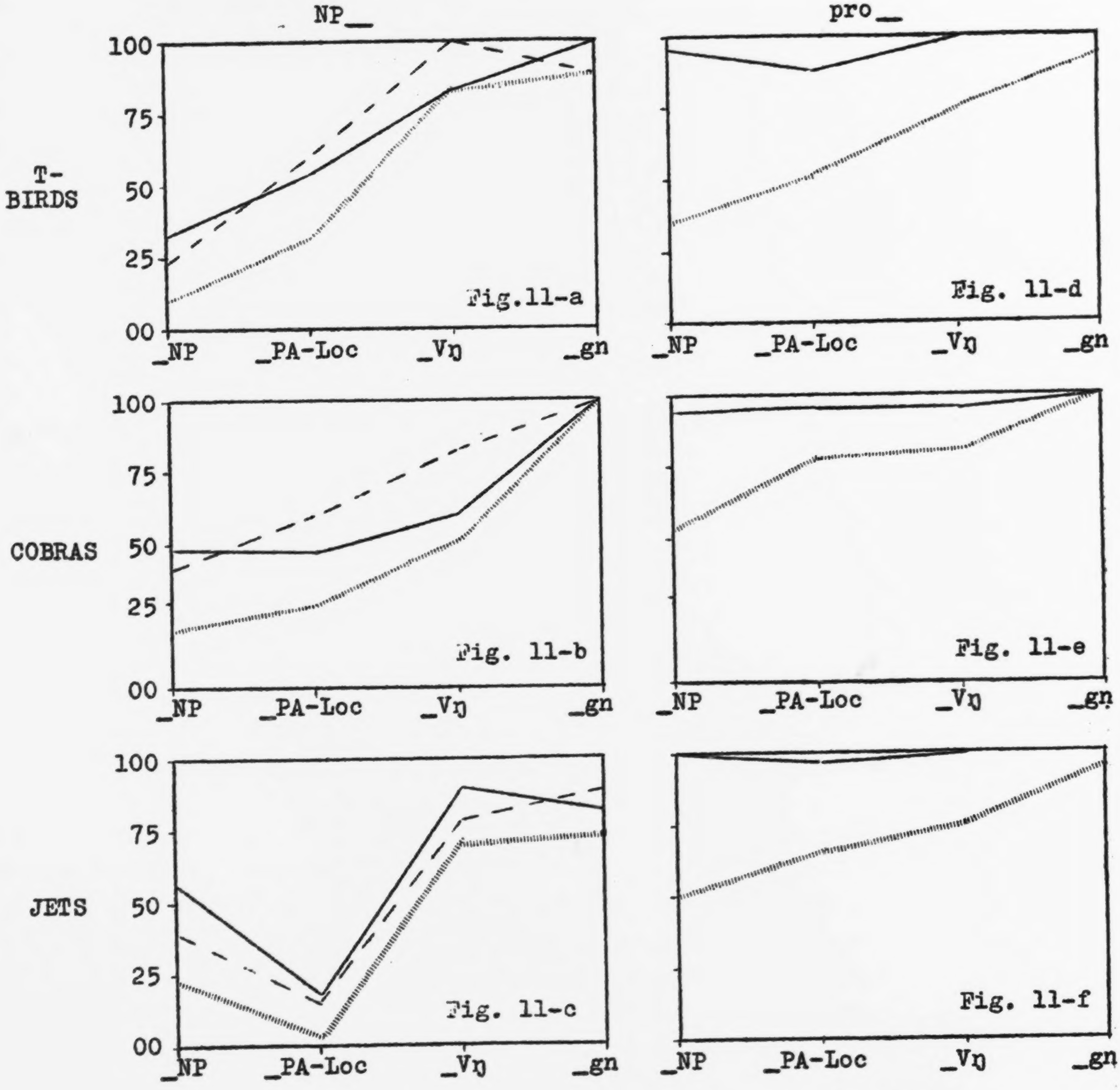
Independence of the preceding and following environments. The grammatical status of the preceding and following elements are only two of the many constraints upon the contraction and deletion rule. We have not yet considered here the effects of the phonological environments. However, before proceding further it is necessary to investigate the relative independence of these two sets of environments. It is possible that one is conditioned by the other--that the effect of a following noun phrase, for example, is entirely different when a pronoun precedes than when another noun phrase precedes. Or going even further, one of these effects could be nothing but the result of unequal distribution of forms in the other environment. For example, a following verb phrase may favor contraction and deletion simply because pronouns occur more frequently before predicates with Verb + ing than they do before predicates with NP.

Figure 5 resolves these questions by displaying the two variable conditions independently. On the left, 5a-c show the effect of the following grammatical category for all sentences with subject noun phrase; on the right, 5d-f show the data for sentences with subject pronouns. Because the total number of forms is considerably reduced for each group (even when single and group styles are combined), the following predicate adjectives and locatives are given together. Still, some of the cells are too small to be reliable, as the table of N at the bottom shows: for the T-Birds, for example, there are only six cases of a following verb after a noun phrase subject, and only eight cases of following gonna, which may be responsible for the irregularity of the pattern at this point.

Figure 5 demonstrates that neither of the environmental constraints--preceding or following--are dependent upon the other, although there is some degree of interaction. There is some degree of irregularity in the patterns with preceding noun phrase: for the Jets, for example, we see that the order of effects of following locative-predicate adjectives vs. following noun phrases is reversed in Figure 5-c. We do not know as yet whether this reversal is constant or reproducible; the data presented here does not exhaust all of the material which is available for the Jets and Cobras, and further analysis will answer such questions.

Figure 5 does show remarkable regularity in the patterns displayed by the three groups, especially in the case of a preceding pronoun. The effect of a preceding pronoun upon contraction is almost a categorical one for all three groups--that is, the contraction rule goes almost to completion--whereas the deletion rule operates variably and regularly across a wide range of frequencies.

Figure 5. Percentages of full, contracted and deleted forms of is according to preceding and following environments



		No. forms			
Fig.	5-a	88	47	6	8
	5-b	71	58	10	13
	5-c	99	61	21	11
	5-d	89	74	35	32
	5-e	49	65	20	31
	5-f	98	88	38	22

Most importantly, all six sections of Figure 5 conform to the model of Figure 4-c--showing that contraction and deletion are governed by similar but slightly different constraints. Contraction and deletion follow the same pattern even when there is a re-ordering in the contraints, as in the NP ~ PA-Loc situation for the Jets in 5-c. With this parallelism, we observe that contraction and deletion have distinct variable inputs and distinct variable constraints which re-apply to deletion after they have applied to contraction. Thus Case 1, in which a contraction rule is followed by a deletion rule, receives ample confirmation. In each case, deletion diverges from contraction on the left and converges on the right. If it is assumed that the deletion rule operates upon the pool of already contracted forms, then the frequency of deletion D/D+C (indicated by a dashed line in Figures 5a-c) regularly rises from left to right (see Table 3 below). In Figures 5d-f it would seem that contraction is virtually independent of the following environment--only traces of variability before noun phrases and predicate adjectives remain. This may be considered the normal result of a variable constraint which has moved to a higher level, producing the semi-categorical pattern shown here.

5. The formal expression of variable rules. The goal of our analysis is to incorporate such variable rules as contraction and deletion into the main body of generative rules needed for a full description of NNE or SE. By absorbing the data of section 4 on systematic variation into the rules, we will be able to resolve questions of ordering and rule form which would otherwise remain undecidable. Furthermore, it will be possible to enlarge our current notion of the "linguistic competence" of a native speaker. To achieve this goal, it is necessary to write single rules for contraction and deletion incorporating the relationships found in Figures 1-5: certain innovations in formal notation will be required which will reflect this enlargement of the concept "rule of grammar".

Linguistic rules are currently conceived in generative grammar as having the general form

(69) $X \rightarrow Y \ / \ A\_\_B$

where X is always re-written as Y in the stated environment, but never re-written as Y otherwise. This is a categorical instruction--the only type of rule which is permitted in any of the traditional approaches to formal grammar.[26] When one is faced with the fact of variation--that the rule does not always apply, it is possible to say that the rule itself is optional--that it may or may not be applied at the discretion of the speaker. We can represent such optionality by writing parentheses around the right-hand member of the rule

(70) $X \rightarrow (Y) \ / \ A\_\_B$

However, if we interpret this notation as meaning no more than the label "optional", it will hardly allow us to embed the facts of systematic variation presented above into the grammar of NNE. The label optional is no more useful in this respect than the label "free variation". It is true that we would come closer to the actual situation in NNE by writing optional contraction and deletion rules rather than obligatory ones. But in so doing, we would be portraying NNE as nothing more than a mixture of random possibilities--a notion quite consistent with the usual conception of "dialect mixture". It is not the object of sociolinguistic analysis to reduce the precision of linguistic rules, nor to add to the vagueness with which linguistic structure is perceived. If the data of the preceding sections is to be utilized in formal rules, it must be shown that the study of variation adds to our knowledge of linguistic structure, and simplifies the situation rather than reducing the precision of the rules by uncontrolled and unaccountable notations.[27] To achieve this end, we associate with each variable rule a specific quantity φ which denotes the proportion of cases in which the rule applies as a part of the rule structure itself. This proportion is the ratio of cases in which the rule actually does apply to the total population of utterances in which the rule can possibly apply, as defined by the specified environment, if it were a categorical rule of the type (69). The quantity φ thus ranges between 0 and 1; for all categorical rules, such as (69), it follows that $\varphi = 1$.

Variable input. It is normally the case that rules do apply categorically, without exception, although there are a great many cases, some of which we consider here, in which some factor interferes with or impedes the full application of the rule so that it is not categorical. It is thus convenient to define φ as

$$(71) \qquad \varphi = 1-k_o$$

where k_o is the variable input to the rule--the factor which limits or constraints the application of the rule. With categorical rules of the type (69), it follows that there is no variable input, and $k_o = 0$, that is, there is no impediment to the operation of the rule. The value of k_o must vary if the variable rule is involved in the process of linguistic change; it is thus a function of the age of the speaker or group. The variable input is also governed by such extra-linguistic factors as contextual style, socio-economic class, sex and ethnic group; we will not be considering such factors here, since our object is the relatively uniform grammar of male adolescent and pre-adolescent Negro members of the vernacular culture in urban ghetto areas.

Variable constraints. The data of section 4 showed that variation in contraction and deletion is governed by a set of constraints such as the effect of a preceding pronoun or a following verb. These variable constraints are features of the environment which are indicated in a variable rule with Greek letters α, β, γ ... as in (72)

(72) $$X \rightarrow (Y) \,/\, \begin{bmatrix} \alpha\text{fea}_i \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \begin{bmatrix} \overline{} \\ \gamma\text{fea}_j \\ \cdot \\ \cdot \end{bmatrix} \begin{bmatrix} \beta\text{fea}_k \\ \cdot \\ \cdot \\ \nu\text{fea}_n \end{bmatrix}$$

Such variable constraints range over + and - just as in the usual generative conventions for variables. However, in conjunction with the parentheses, given the automatic reading that

(73) $$\varphi = 1 - (k_o - \alpha k_1 - \beta k_2 \; . \; . \; . \; k_n)$$

where $k_o..n$ are constants which can be determined by empirical studies. These conventions are so designed that if the variable feature is present or + in a given sub-set of sentences, it favors the application of the rule. Thus if α in (72) is +, k_1 in (73) is subtracted from the variable input k_o, there is less impediment to the operation of the rule, and φ is larger. Since φ does not apply to individual sentences, but rather to sets of sentences, we here designate that sub-set of the total population of utterances defined by the rule in which α is + and $[\text{fea}_i]$ is present as $\varphi(\alpha)$. The complementary subset in which α is -, and $[\text{fea}_i]$ is absent, is designated $\varphi(\sim\alpha)$. Thus in general, the use of variable constraints indicates that

(74) $$\varphi(\alpha) > \varphi(\sim\alpha).$$

The invariance condition. In the usual notation for categorical rules, the environment / __ $[+\text{fea}_i]$ means that the rule always applies for that subset of sentences in which $[\text{fea}_i]$ occurs in that position, and never applies for the subset where $[\text{fea}_i]$ does not occur. In other words, $\varphi(\text{fea}_i) = 1$, $\varphi(\sim\text{fea}_i) = 0$. For variable rules, the notation still allows us to register the fact that the rule never applies in certain cases. Thus, if the environment includes / __ [+cons], then the rule never applies for the subset of sentences in which [-cons] follows the item in question. Thus if our contraction rule reads $/ \;.\;.\begin{bmatrix}\underline{\quad\quad} \\ +T\end{bmatrix}.\;.$ it follows that contraction never occurs for the set of sentences in which the verb is not finite, associated with [-T]. But on the other hand, we are still lacking a means of incorporating into our contraction rule the fact that when a nasal consonant follows the schwa, contraction to I'm is for all practical purposes universal; that is, in the presence of a given feature, a variable rule becomes categorical. We need a formal means, then, of expressing the feature of invariance in a variable rule. The symbol * is used to designate such an invariant feature as in (75):

$$(75) \qquad X \rightarrow (Y) \;/ \begin{bmatrix} \alpha fea_i \\ \cdot \\ \cdot \\ \cdot \\ {*}fea_x \end{bmatrix} \begin{bmatrix} \overline{\quad\quad} \\ \gamma fea_k \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \begin{bmatrix} \beta fea_j \\ \cdot \\ \cdot \\ \cdot \\ \nu fea_n \end{bmatrix}$$

which is automatically read as (76):

$$(76) \qquad \varphi = 1 - \left(\frac{-1*1}{-2}\right)(k_o - \alpha k_1 - \beta k_2 \;.\;.\;.\; \nu k_n)$$

Thus for sentences in which [fea_x] occurs as +, the invariance factor is $\frac{-1-(+1)}{-2} = 0$, the entire variability factor goes to zero, and $\varphi = 1$. But where [fea_x] is -, the invariance factor is $\frac{-1-(-1)}{-2} = 1$, and the value of φ is unaffected. Thus the expression $\frac{-1*1}{-2}$ is a device for converting +,- values into 0,1 values: it is the formal equivalent of the statement that * converts a variable rule into a categorical one. More generally, we can state that the symbol * has the property that for all rules

$$(77) \qquad \varphi(*) = 1; \qquad \varphi(\sim *) = \varphi$$

The effect of the various notations on values of φ can be seen in the following array:[27a]

(78)

Rule contains the notation ...	$\varphi(fea_i)$	$\varphi(\sim fea_i)$
$+fea_i$	φ	0
$-fea_i$	0	φ
αfea_i	$1-(k_o-k_1\ldots)$	$1-(k_o+k_1\ldots)$
$-\alpha fea_i$	$1-(k_o+k_1\ldots)$	$1-(k_o-k_1\ldots)$
$*fea_i$	1	φ
$-*fea_i$	φ	1

Ordering of variable constraints. The order of the Greek letters α, β, γ is not arbitrary in these conventions; in any rule of the form (75), with automatic reading (76), it follows that

(79) $$k_1 > k_2 > k_3 > \ldots k_{n-1} > k_n$$

The values of these constants can be determined, within certain limits, by data such as that presented in section 4. But the question must be raised, what is linguistically significant in these data? It is unlikely that it will be important for us to know that the copula is deleted 82°/o of the time by Speaker A and 79°/o of the time by Speaker B. The structures we are examining are not a series of numbers, but rather a series of relationships--between the environment and the /z/, and between one environmental constraint and another. The constraints of a preceding pronoun and a following noun phrase are not equivalent: they appear to be ordered. This ordering is most apparent in the relationships of the cross-products, where one feature is favorable and the other unfavorable. If no statements could be made about the relationships of such cross-products, then it would be apparent that we have a very weak type of ordering; a strong statement would be that all of the cross-products are strictly ordered. We can formalize this *postulate of* geometric ordering as follows:

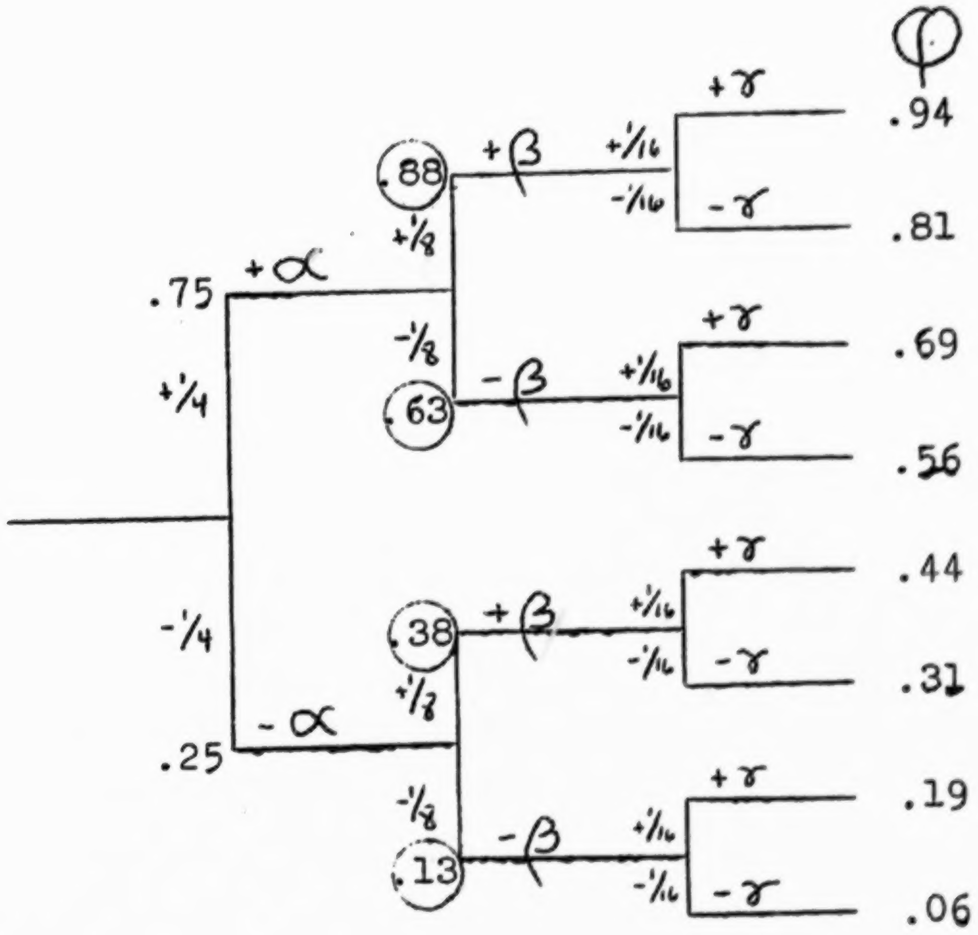
(80) If $\chi_1, \chi_2, \ldots \chi_n$ are variable constraints upon a rule r, then for any given $\chi_1, \chi_2, \ldots \chi_{i-1}$, $\varphi_r(\chi_i) > \varphi(\sim \chi_i)$.

In other words, each constraint in the hierarchy outweighs the effects of all constraints below it. If we take sentences with α fixed, then any sub-set of these with β as + will show the rule applying in a higher proportion of cases than any sub-set with β as -. The cross-product with β as + and $\gamma, \delta\ldots$ as all - will still show a higher value of φ than the cross-product with β as - and all lower constraints as +.[28]

We can generate such a set of ordered cross-products by arbitrarily assigning the values $k_o = 1/2$, $k_1 = 1/4$, $k_2 = 1/8\ldots n$. This series may be displayed as a tree, as shown in Figure 6.[29]

As section 4 shows, the relations symbolized by $\alpha, \beta, \gamma\ldots$ are quite binding, and the data for each individual shows that they hold for very small numbers. But the higher order relationships which concern the ordering of these constraints within the hierarchy are not so uniform. Although the major constraints hold for all groups, there is variation from one group to another in the effect of a following noun phrase as compared to a following adjective. Furthermore, the phonological constraint of the effect of a following vowel or consonant (not discussed in this paper) is a marginal or inconsequential effect for younger groups, and gradually assumes more importance with age. There is reason to believe that changes in the hierarchy of constraints represent a basic mechanism of linguistic development--as it affects a whole community in the course of linguistic evolution, or as it affects peer groups in regular age-grading.[30] This discussion, however, will be confined to the major relations of order within a relatively uniform grammar. For this purpose, one more variable constraint upon contraction

Fig. 6. Geometric ordering of variable constraints α, β, γ.


φ
.75
+α
+1/4
-1/4
-α
.25
.88
+1/8
-1/8
.63
.38
+1/8
-1/8
.13
+β
-β
+β
-β
+1/16
-1/16
+1/16
-1/16
+1/16
-1/16
+1/16
-1/16
+γ
-γ
+γ
-γ
+γ
-γ
+γ
-γ
.94
.81
.69
.56
.44
.31
.19
.06

and deletion must be presented: the effect of a preceding vowel as against a preceding consonant.

6. *The effect of a preceding vowel on contraction and deletion*. There are a number of phonological constraints upon the operation of contraction and deletion, but the most important, from the standpoint of magnitude and linguistic significance, is whether or not the preceding element ends with a consonant or a vowel. Most subject pronouns end with stressed vowels[31], but other noun phrases can be sub-classified in many ways according to their final segments. The most useful sub-categories of the environments for the contraction and deletion of *is*, are as follows:

(a)	-S__	After noun phrases ending in sibilants.
(b)	$-K^{\circ}$__	After noun phrases ending in non-sibilant voiceless consonants.
(c)	$-K^{v}$__	After noun phrases ending in non-sibilant voiced consonants.
(d)	-V__	After noun phrases ending in vowels[32].

It is no accident that the first three of these categories are the same as those used to describe forms of the English /Z/ morpheme[33]. But whereas the usual rules can treat categories (c) and (d) as one (the "elsewhere" or "other voiced segment" category), the distinction between (c) and (d) will be critical in the analysis of contraction and deletion.

Table 3 shows the percentages of full, contracted and deleted forms for all six groups studied in section 4 according to the phonetic form of the preceding element. Examining the percentages of full forms, we can immediately say that

(1) In all cases, there are fewest full forms after pronouns; contraction is, therefore, almost categorical after pronouns, as observed in

TABLE 3

PERCENTAGES OF FULL, CONTRACTED, AND DELETED FORMS ACCORDING TO PHONETIC FORM OF PRECEDING ELEMENT FOR SIX GROUPS IN SINGLE AND GROUP STYLES COMBINED

	$-K^{o}$__	$-K^{v}$__	-S__	-V__	pro__
Thunderbirds					
Full	83	70	62	43	05
Contracted	05	28	00	30	42
Deleted	12	02	38	27	53
	100	100	100	100	100
N:	24	92	21	79	255
Cobras					
Full	54	58	67	10	03
Contracted	08	09	06	53	28
Deleted	38	33	27	37	69
	100	100	100	100	100
N:	13	33	18	32	136
Jets					
Full	89	58	80	42	00
Contracted	00	14	00	45	39
Deleted	11	28	20	13	61
	100	100	100	100	100
N:	28	65	29	69	269
Oscar Brothers					
Full	93	71	68	40	04
Contracted	00	21	12	40	54
Deleted	07	08	20	20	42
	100	100	100	100	100
N:	15	14	41	37	95
Working-class adults					
Full	75	69	88	45	39
Contracted	08	21	03	45	47
Deleted	16	10	09	10	14
	99	100	100	100	100
N:	48	100	75	83	200
Inwood groups					
Full	42	30	97	13	00
Contracted	58	70	03	87	100
Deleted	00	00	00	00	00
	100	100	100	100	100
N:	12	46	34	65	61

section 4 above.

(2) In all cases, there are fewer full forms after noun phrases ending in vowels than after those ending in consonants, but more than after pronouns. In other words, the fact that pronouns end in vowels accounts for some, but by no means all, of their effects upon contraction.

(3) In all cases but one[34], there is a small, but distinct tendency for there to be more full forms after voiceless consonants than voiced.

(4) There are almost no contracted forms after sibilants, although a few definitely can be observed, contrary to the usual conception. But quite a few forms of is have apparently undergone both contraction and deletion: if we consider that forms such as The fish is... follow the same rules as the rest of the other NNE sentences, then it appears that deletion is practically categorical after sibilants.

Since noun phrases are relatively sparse as compared to subject pronouns, the numbers for all of these sub-categories are not large enough for us to study the operation of deletion within them. Table 4 therefore compares the operations of contraction and deletion by combining $-K^{\circ}\_\_$ and $-K^{v}\_\_$ into a single category $-K\_\_$.

The contraction rule is seen as having operated upon full forms to produce the contracted and deleted forms; and deletion as then operating upon the resulting pool of contracted forms.

(82) $\varphi_C = \frac{C+D}{F+C+D}$; $\varphi_D = \frac{D}{C+D}$

TABLE 4

FREQUENCY OF OPERATION OF DELETION AND CONTRACTION RULES WITH PRECEDING CONSONANT OR VOWEL FOR SIX NNE GROUPS IN SINGLE AND GROUP STYLES COMBINED

	φ_C $\frac{C+D}{F+D+C}$	N	φ_D $\frac{D}{C+D}$	N
Thunderbirds				
-K__	.28	116	.16	32
-V__	.57	79	.47	45
pro__	.95	255	.56	241
Cobras				
-K__	.41	46	.80	20
-V__	.90	32	.41	29
pro__	.97	136	.71	132
Jets				
-K__	.32	93	.70	30
-V__	.58	69	.22	40
pro__	1.00	269	.61	269
Oscar Brothers				
-K__	.17	29	(.40)	5
-V__	.59	37	.33	22
pro__	.96	95	.44	91
Working-class adults				
-K__	.30	148	.38	59
-V__	.55	83	.18	46
pro__	.61	200	.77	99
Inwood groups				
-K__	.67	58	.00	39
-V__	.87	65	.00	60
pro__	.99	142	.00	141

For the Cobras, Jets, Oscar Brothers and adults, it appears that a preceding vowel favors contraction, while exactly the opposite situation prevails with deletion: the rule applies more frequently when a consonant precedes. Figure 7 shows the striking character of this reversal, and runs counter to the parallelism of contraction and deletion which has prevailed up to this point. The Inwood group shows no deletion, but we observe that contraction is also favored by a preceding vowel in their case. Only the youngest group, the Thunderbirds, does not show this effect: for them, a preceding vowel favors both contraction and deletion.[35]

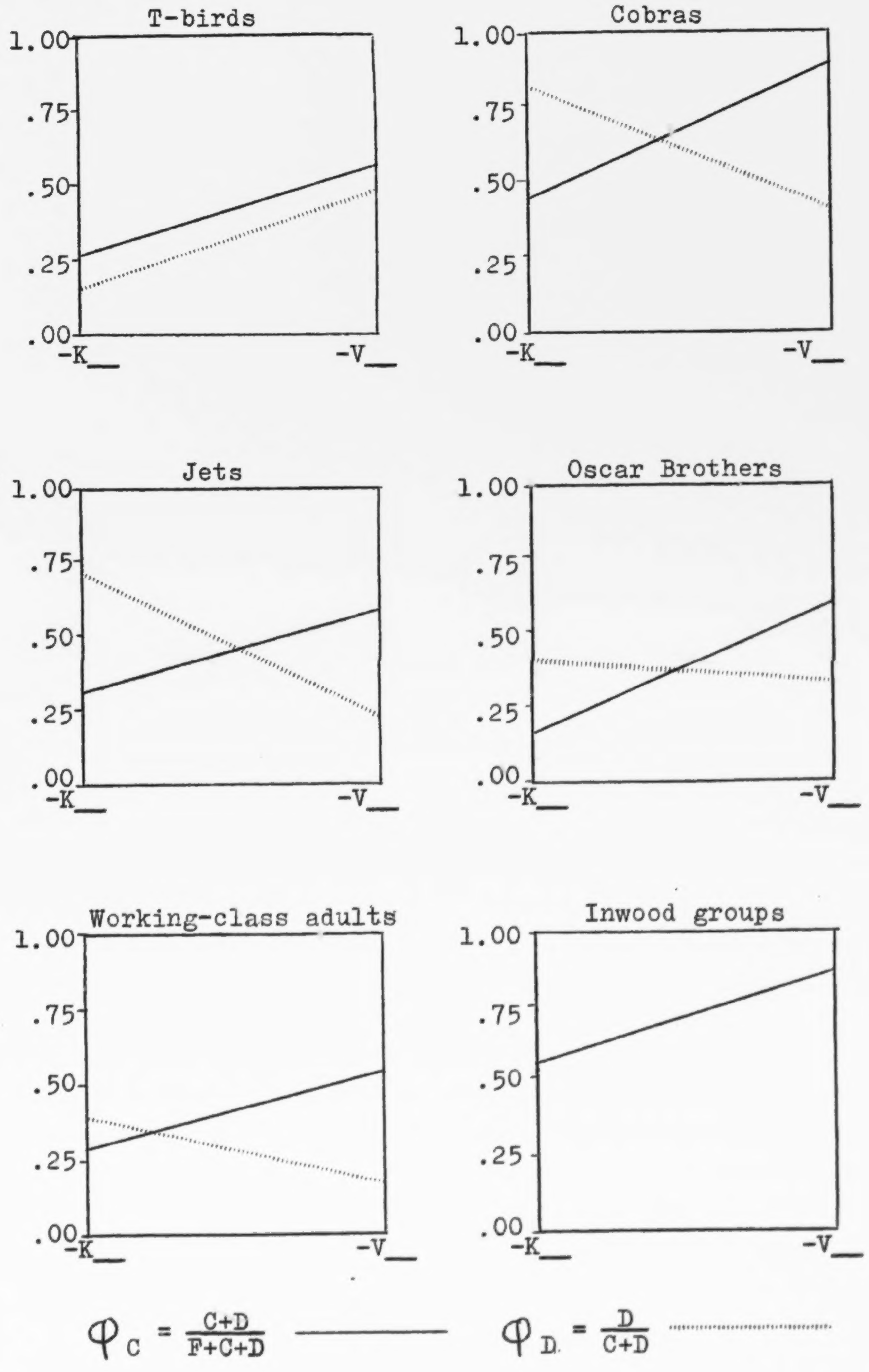
The pattern which prevails can be illustrated by (83) and (84).

(83) Stanley is here. $\overset{C}{\Rightarrow}$ Stanley's here. $\overset{D}{\rightarrow}$ Stanley here.
CV VC CVC — CVC CVC — CV CVC

(84) Stan is here. $\overset{C}{\rightarrow}$ Stan's here. $\overset{D}{\Rightarrow}$ Stan here.
CVC VC CVC — CVCC CVC — CVC CVC

In the case of a subject noun ending in a vowel, we see that contraction acts to reduce a CVVC sequence to CVC. (It is true that the first vowel may be diphthongized so that a glide interposes between the two vowels in the actual phonetic output, but this is not always the case in NNE.) On the other hand, when contraction operates upon a subject noun ending in a consonant, the result is a consonant cluster. There are a number of rules operating throughout NNE which reduce consonant clusters, although there is no single rule for all cases. In general, it can be said that NNE, like English and most Indo-European languages, disfavors final consonant clusters, and there are many examples of historical processes operating to reduce them. This tendency runs strongly in NNE, though it is by no means extreme in this respect.[36] In any case, the way in which contraction and deletion are opposed with respect to the preceding vowel clearly

Fig. 7. Effect of a preceding consonant or vowel upon operation of the contraction and deletion rules for six groups: single and group styles combined



demonstrates that both contraction and deletion are phonological processes; furthermore, our original analysis that deletion is the removal of a lone consonant produced by contraction receives strong confirmation from the data presented here.[37]

It is also apparent from Table 4 that the effect of a preceding pronoun upon contraction and deletion is in part dependent upon, but in part distinct from, the effect of a preceding vowel. Almost all pronouns end in tense vowels, and it is plain that contraction is heavily favored when the subject is a pronoun. But the effect is much stronger than for other noun phrases ending in vowels--in fact, it is to all effects a categorical rather than a variable rule. In the contraction rule, there will therefore be an entry /[*pro]__ which states that after pronouns, the rule is not a variable but a categorical one. In the case of deletion, it can be seen that the rule operates much more often when a pronoun precedes than when another noun phrase ending in a vowel precedes. Therefore the effect of a preceding pronoun will be one of the variable constraints upon deletion, though not necessarily the primary one.

Independence of phonological and grammatical constraints. To this point, we cannot be sure that the effect of a preceding vowel or consonant is not the product of some odd distribution of noun phrases before various complement categories, since the data of Tables 3 and 4 treats all such categories alike. As we have seen in Table 2, a following verb strongly favors both contraction and deletion, and it is possible that the noun phrases which precede verbs are different from those which precede predicates. Table 5 shows the percentages of contraction and deletion, on the same basis as Table 4, but with the proportions for four following grammatical categories shown separately. Since the numbers necessarily become quite small, the figures for the four adolescent NNE groups are grouped together: the T-Birds, the Cobras, the Jets and the Oscar Brothers. The result shows that the opposing effect of a preceding vowel and consonant

TABLE 5

FREQUENCY OF OPERATION OF DELETION AND CONTRACTION RULES ACCORDING TO PRECEDING AND FOLLOWING ENVIRONMENTS FOR FOUR ADOLESCENT NNE GROUPS IN GROUP STYLE ONLY

	__NP				__PA/LOC			
	φ_C	N	φ_D	N	φ_C	N	φ_D	N
-K__	.37	35	.62	13	.25	32	.50	8
-V__	.80	64	.29	51	.70	23	.37	16
pro__	.94	32	.40	30	.98	65	.56	64

	__Vb				__gonna			
	φ_C	N	φ_D	N	φ_C	N	φ_D	N
-K__	.65	4	1.00	9	.89	9	.87	8
-V__	.86	14	.33	12	1.00	6	1.00	6
pro__	.97	34	.79	33	1.00	23	.96	23

holds for all syntactic environments, except in the case of a following future in gonna, where both contraction and deletion are close to categorical, and the numbers are very small. In the other cases, we again observe that the effect of a preceding pronoun is semi-categorical for contraction, and that deletion is much stronger with a preceding pronoun than with a noun ending in a vowel. Table 5 thus provides us with additional confirmation of our analysis of the relations between contraction and deletion.

7. The rules for contraction and deletion. We can now incorporate the quantitative data of section 4 and 6 into the logical development of ordered rules for contraction and deletion of section 1-3, using the formal apparatus of section 5. The outline on the following page shows a series of sixteen phonological rules of NNE in which the contraction rule (9) and the deletion rule (13) for is are embedded. The contraction and deletion rules are given in full; other rules are shown in enough detail to illustrate their general character and their relation to (9) and (13).

Only a few of these rules are peculiar to NNE; half of them are part of the basic machinery of SE, and operate in exactly the same fashion in NNE. This is the case for the eight rules marked with **. The nuclear stress rule operates well before any of the others to provide conditions for vowel reduction, as discussed above; the weak word rule (4) and vowel reduction (5) provide the [ə] upon which rule (9) operates. Rules (2,3,6,7,8) are relevant to other contractable items such as have, has, will and are, and will be considered briefly below. Rules (10) and (11) are concerned with -sps, -sts, -sks, and -t,d clusters in general, which intersect with the grammatical category of the past tense, and are considered in some detail elsewhere. Once we establish the basic conditions for contraction by rules (0),(4), (5), the behavior of is is governed by the five rules (9), (12) (13), (14)

SIXTEEN PHONOLOGICAL RULES OF NNE

**(0)	Nuclear stress rule	$V \rightarrow \overset{\prime}{V} / [\overset{\prime}{V}]\ldots[\,'\,]\,]_{\alpha}$
	⋮	
**(1)	Centralization of vowels before r	$V \rightarrow (\emptyset) / \begin{bmatrix} \underline{} \\ -\text{low} \end{bmatrix} r\,[\alpha\text{cons}]$
(2)	Vocalization of r	$r \rightarrow (\text{ə}) / [-\text{cons}]\_\_\,\alpha(\#\#)\ {*}(-V)$
(3)	Vocalization of l	$l \rightarrow (\overset{\text{ł}}{\frown}) / [-\text{cons}]\_\_(\#\#)\ -V \ldots$
**(4)	Weak word rule	$[3\text{str}] \rightarrow [-\text{str}] / [\_\_, +W]$
**(5)	Vowel reduction	$V \rightarrow \text{ə} / [\_\_, -\text{str}, -\text{tense}]$
(6)	Loss of postvocalic ə	$\text{ə} \rightarrow (\emptyset) / [+\text{voc}, -\text{cons}, \alpha\text{high}]\_\_\#\# \ldots$
(7)	Loss of postvocalic $\overset{\text{ł}}{\frown}$	$\overset{\text{ł}}{\frown} \rightarrow (\emptyset) / [+\text{voc}, -\text{cons}\ldots]\_\_\#\# \ldots$
**(8)	Loss of initial h	$h \rightarrow (\emptyset) / \_\_\text{ə}\ C_0^1\ \#\#$
**(9)	Contraction	$\boxed{\text{ə} \rightarrow (\emptyset) / \begin{bmatrix} {*}\text{pro} \\ \alpha V \end{bmatrix} \#\# \begin{bmatrix} \underline{} \\ +T \end{bmatrix} C_0^1\ \#\# \begin{bmatrix} \alpha\text{Vb} \\ \beta\text{gn} \\ ?\text{NP} \end{bmatrix}}$
(10)	*Simplification of -sK clusters*	$[-\text{cont}] \rightarrow (\emptyset) / [+\text{strid}]\_\_\#(\#) \begin{bmatrix} {*}\text{strid} \\ \alpha(-V) \end{bmatrix}$
(11)	General simplification of -t,d clusters	$t,d \rightarrow (\emptyset) / \begin{bmatrix} \alpha\text{cons} \\ \vdots \end{bmatrix} \gamma(\#)\_\_\ \beta(-V)$
(12)	Assibilation of -t	$t \rightarrow s / [\_\_, +\text{pro}]\#(\#)[+\text{strid}]\#\#$
(13)	Deletion	$\boxed{[+\text{cont}] \rightarrow (\emptyset) / \begin{bmatrix} -\beta V \\ \gamma\,\text{pro} \\ {*}\text{strid} \end{bmatrix} \#\# \begin{bmatrix} \underline{} \\ -\text{nas} \end{bmatrix} \#\# \begin{bmatrix} \alpha\text{Vb} \\ \beta\text{gn} \\ ?\text{NP} \end{bmatrix}}$
**(14)	Epenthetic vowel	$\emptyset \rightarrow \text{ə} / [+\text{strid}]\#(\#)\_\_[+\text{cont}]\#\#$
**(15)	Voicing assimilation	$[-\text{voc}] \rightarrow [\alpha\text{voice}] / [\alpha\text{voice}]\#(\#)\_\_\#\#$

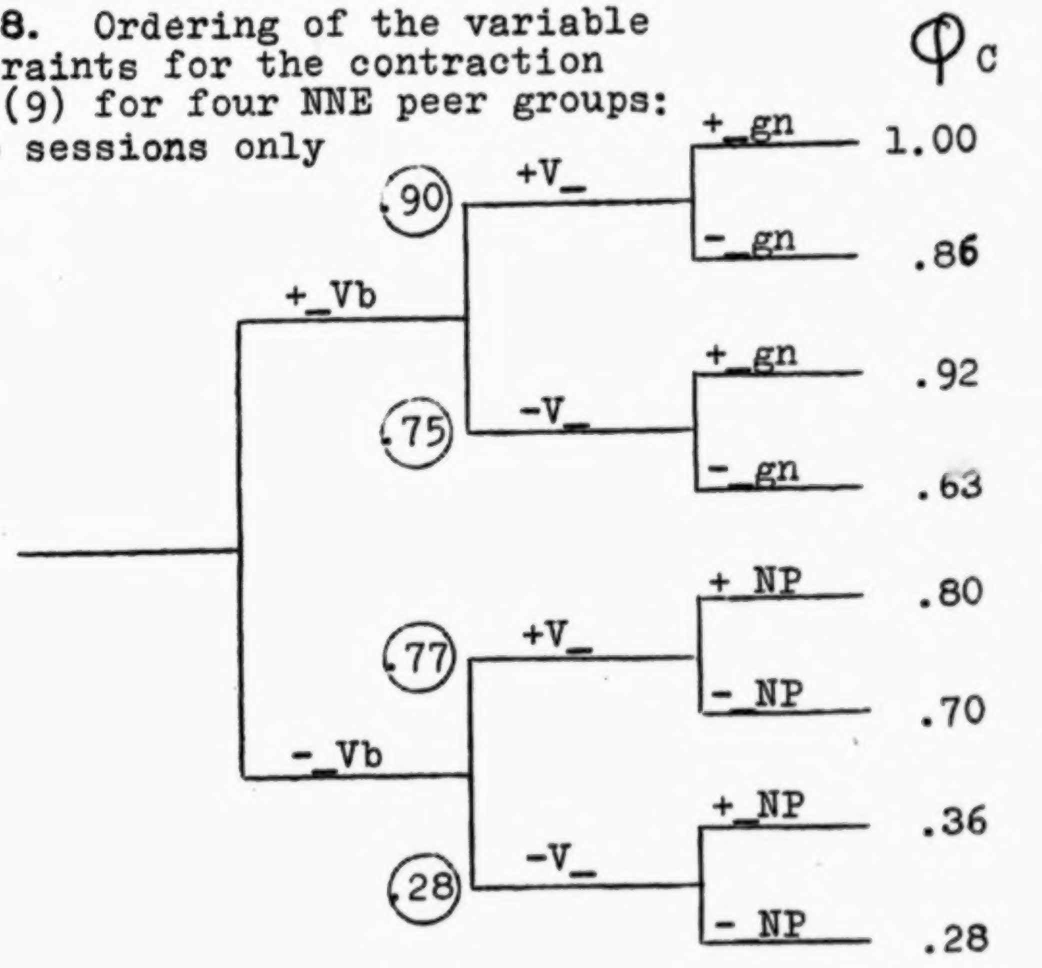
and (15), which we will examine here.

Form of the contraction and deletion rules. Rule (9) appears as the removal of a schwa, occurring initially before a single consonant, in a word with the tense-marker incorporated. When a pronoun proceeds in NNE, the rule is (semi-)categorical, as indicated by the invariance condition *. The variable constraints do not show a high degree of order: a preceding vowel and a following verb have approximately equal effect in promoting the application of the rule, while the effect of a following future in *gonna* is somewhat less. Figure 8 shows the resulting tree, incorporating data from the four vernacular NNE groups in group interaction. There are two α variables, since __Vb and V__ are equivalent. Among the various non-verbal predicates, the effect of a following noun phrase as against a following predicate adjective or locative, is indicated clearly enough in the total results, but it is not consistent enough among the various peer groups to warrant incorporating it into the general rule for NNE.

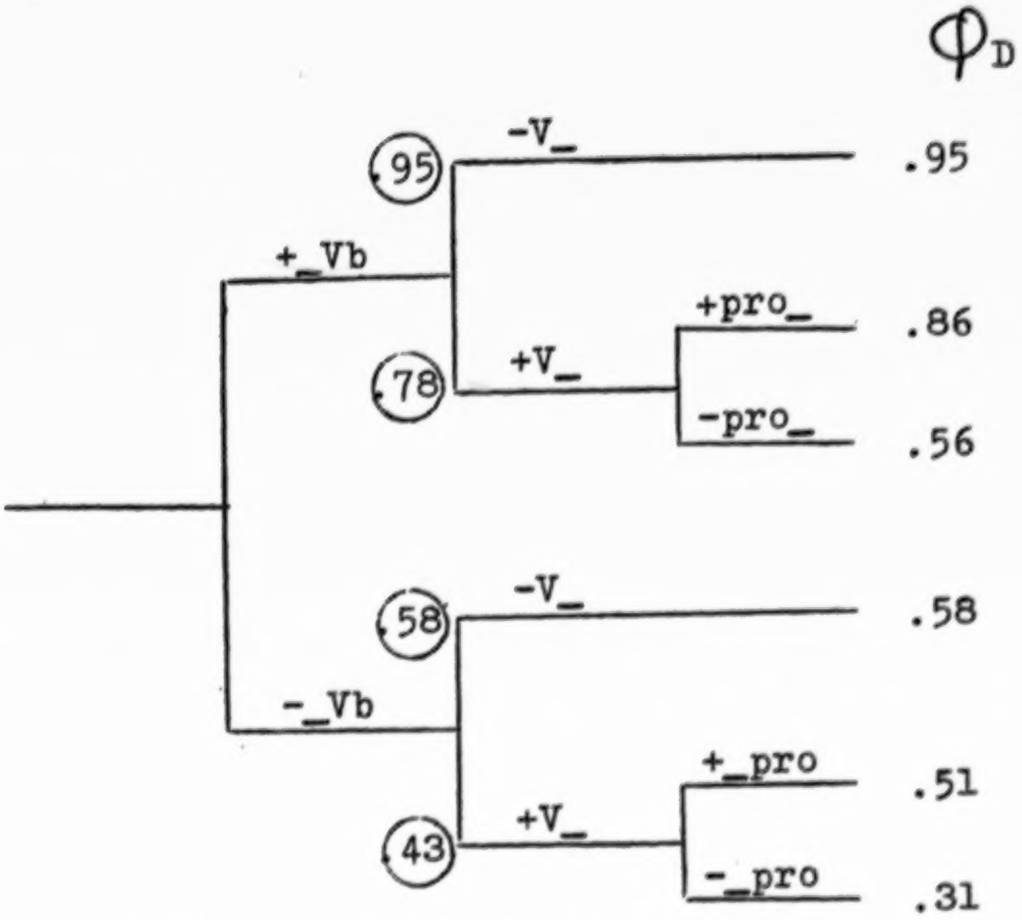
The deletion rule (13) appears as the removal of a lone oral continuant between word boundaries. Here the variable constraints show a higher degree of order, as indicated in Figure 9. The primary constraint is the effect of a following verb, and the secondary constraint the effect of a preceding vowel--but reversing the polarity for the contraction rule. The combination of these two yields the series of values .95 - .78 - .58 - .43 which shows geometric ordering with an input value at a higher level than that shown in Figure 6. The third effect, that of a preceding pronoun is almost well ordered, but of course is not represented on the V__ branches. The __gonna constraint is not shown here, but has about the same weight as V__, and like all other variables except V__, it follows the same direction as with contraction.

The quantitative data presented in this paper is sufficient to establish the major variable constraints upon these rules--constraints which are independent of each other and which re-cur

Fig. 8. Ordering of the variable constraints for the contraction rule (9) for four NNE peer groups: group sessions only



ϕ_C
.90
.75
.77
.28
+_Vb
-_Vb
+V_
-V_
+V_
-V_
+_gn
-_gn
+_gn
-_gn
+_NP
-_NP
+_NP
-_NP
1.00
.86
.92
.63
.80
.70
.36
.28



ϕ_D
.95
.78
.58
.43
+_Vb
-_Vb
-V_
+V_
-V_
+V_
+pro_
-pro_
+_pro
-_pro
.95
.86
.56
.58
.51
.31

Fig.3-14. Ordering of the variable constraints for the deletion rule (13) for four NNE peer groups: group sessions only

regularly in almost all styles and peer groups. It will no doubt be possible to modify this presentation in the future, as more data is accumulated; there are many interesting questions concerning the role of various predicate types to be investigated. But the purpose of this type of analysis is not to explore every conceivable constraint upon a variable rule to the limits of reproducibility, but rather to apply the logic of these converging (and diverging) patterns to establishing the place, form and order of the deletion and contraction rules of NNE.

One of the first, and most obvious arguments for order springs from the predominance of i's, tha's and wha's [ɪs, ðæs, wʌs] as the NNE phonetic output of underlying it is, that is and what is. At first glance it seems obvious that the assimilation of the /z/ to the preceding voiceless stop has produced an [s] which is not subject to the deletion rule, and therefore deletion does not apply.[39] In the light of this evidence, we would order the voicing assimilation rule before the deletion rule. We would then have derivations such as the following:

(85)

ɪt##ɪz	
ɪt##əz	vowel reduction
ɪt## z	contraction
ɪt## s	voicing assimilation
	[deletion--does not apply]
ɪs## s	assibilation
ɪ ## s	reduction of geminates

After a sibilant, we have two possible routes, as shown in (86)

(86)

A	B	
fɪš##ɪz	fɪš##ɪz	
fɪš##əz	fɪš##əz	vowel reduction
	fɪš## z	contraction
	fɪš## s	voicing assimilation
	fɪš##	deletion

The first tendency is to deny that contraction can take place after sibilants, though we do encounter a few cases. But the existence of a sizeable number of zero forms makes it seem clear that route B is followed. Deletion of /z/ after a sibilant must therefore be categorical, as indicated in the rule by [*strid__].

However, the case of the plural fishes poses a more difficult problem:

(87)	fɪš#z	
	fɪš#s	voicing assimilation
		deletion [does not apply across inflectional boundary]
	*fɪš#əs	epenthesis

This result is plainly wrong, and we are forced to conclude that voicing assimilation is ordered after epenthesis, so that it will not assimilate /z/ to a preceding voiceless sibilant. But epenthesis must come after deletion, for the whole force of the evidence in sections 4 and 6 indicates that deletion is the removal of a lone consonant; we do not find any remnants of an epenthetic vowel in expressions such as *That des' [ə] mine or *One fish [ə] on my line.[40] And assibilation must precede deletion if forms such as i's are to survive as regularly as they do. Therefore the correct order must be

contraction
assibilation
deletion
epenthesis
voicing assimilation

It is an attractive notion to place the rule of voicing assimilation last, since this is actually a very general constraint upon the form of final clusters which contain morpheme boundaries. But this order is contrary to the notion expressed above that in i's, /z/ is assimilated to [s] before deletion. The contradiction lies in the assumption that the [s] of [ɪs] is derived from is, as indicated by the practice in dialect literature of writing i's. It now seems clear that this [s] is the assibilated [t] of it--the verb is has entirely disappeared, leaving only this footprint on the preceding pronoun, in the following fashion:

(88)	ɪt##ɪz		
	ɪt##əz	vowel reduction	(5)
	ɪt## z	contraction	(9)
	ɪs## z	assibilation	(12)

We have already seen that deletion must be categorical after sibilants, so it follows that the result is

ɪs## deletion (13)

The order (12)-(13)-(14)-(15) as shown in the rules therefore gives the correct results. Rule (12) shows that assibilation is restricted to words with [+pro]; there are four such pronouns ending in -*t*: *it*, *that*, *what* and *lot*. It is a rule which applies with a somewhat lower input for other [WNS] dialects of English. Neither NNE nor WNS use [pæsgʊd] for *Pat's good*, nor rhyme with [ðæsgʊd] for *That's good*. But it is possible that the restriction of the assibilation rule to pronouns and lone /z/ is too sharp: the rule may apply to other frequent forms ending in -*t*, such as *outside*. However, we do not have enough evidence at present to judge whether the rule operates regularly in cases such as these, and intuitions are quite unreliable in these areas of morphological condensation.

Given the rule order shown above, we have the derivations

(89)	A	B	C		
	fish is	*fish is*	*fish [pl]*		
	fɪš##ɪz	fɪš##ɪz	fɪš#z		
	fɪš##əz	fɪš##əz		vowel reduction	(5)
	fɪš## z			contraction	(9)
	fɪš##			deletion	(13)
			fɪš#əz	epenthesis	(14)
				voicing assim'n	(15)

The form *fish is* can follow route A or B, depending on whether contraction applies, yielding The *fish good today* or *The fish is good today*. The plural fishes appears only as [fɪš#əz], since deletion does not apply across an inflectional boundary. The epenthesis rule can also apply to [fɪš##z], so that we could have the alternative derivation to yield the same result as B

D

fish is

(90)	fɪš##ɪz		
	fɪš##əz	vowel reduction	(5)
	fɪš## z	contraction	(9)
		deletion	(13)
	fɪš##əz	epenthesis	(14)

In this case, the deletion rule would not apply categorically after sibilants. However, the quantitative evidence of Table 4 shows that derivation A is heavily favored, and if the contraction rule applies with roughly the same frequency after sibilants as after other consonants, it seems that deletion is (semi-)categorical after sibilants, yielding very rarely a contracted but undeleted form [fɪšs]. (The operation of the epenthesis and voicing assimilation rules would normally yield a result identical with the full, uncontracted form; such forms as [fɪšs] are doubly rare since the normally categorical epenthesis rule must also be suspended in such cases.)

One prominent characteristic of NNE morphology is that final clusters in -sts, -sps, and -sks are obligatorily simplified, so that an underlying form //test// (which shows up in the verb form testing) cannot have a plural [tɛsts]. The phonetic form which does appear is chiefly [tɛsəz]. This form is derived by the following sequence:

(91)	tɛst#z		
	tɛs #z	simplification of -sC clusters	(10)
	tɛs #əz	epenthesis	(14)

In this environment, the simplification of -sC clusters is categorical, as indicated in rule (10). For the sequence in The test is... one can obtain:

(92)	tɛst##ɪz		
	tɛst##əz	vowel reduction	(5)
	tɛst## z	contraction	(9)
	tɛs ## z	simplification	(10)
	tɛs ##	deletion	(13)

But the contraction rule is not categorical here; when it does not apply, the simplification of -sC clusters now takes place before a following vowel, and it is possible to get either A or B:

(93)

	A	B		
	tεst##ɪz	tεst##ɪz		
	tεst##əz	tεst##əz	vowel reduction	(5)
			contraction	(9)
	tεs ##əz		simplification	(10)
			deletion	(13)

It appears then that rules (5-15) are strictly ordered, with the exception that the general -t,d simplification rule (11) cannot be ordered with respect to the deletion rule (13), since they apply across different boundaries, nor with respect to the assibilation rule (12), which never applies to clusters.[41]

Other contractable verbs. Rules (2), (3), (6) and (7) operate upon liquids /r/ and /l/, as general phonological rules of NNE, and affect other verb forms that are later contracted and deleted--chiefly are and will. The vocalization of these consonants is a process which occurs in somewhat different form in many other English dialects, but the loss of the resulting vocalic glide by rules (6) and (7) is quite peculair to NNE. Thus rule (2) in its full form for NNE is categorical for final and pre-consonantal r and variable only in pre-vocalic position.

(93) $$\begin{bmatrix}+\text{voc}\\+\text{cen}\end{bmatrix} \rightarrow ([-\text{cons}]) \ / \ [-\text{cons}] \ \_\_\_ \ \alpha(\#\#) \ {*}({\sim}V)$$

while the corresponding rule for the WNS vernacular of New York City is variable where NNE is categorical.[42] In studying these vocalization processes, it becomes evident that they represent the sudden or gradual loss of a single feature--[+consonantal] gives way to [-consonantal]. It is therefore essential that weakly constricted, "humped" [r] and [ə] should differ only in that one feature. These two segments are shown here as sharing the feature [+central] which differentiates [r] and [ə] from [l] and the back lateral glide [ɫ̯]. The glides themselves are removed by variable rules (6) and (7) when they follow vowels, producing the well-known lower prestige Southern forms *po'* [po] and *do'* [do] for [pɔ^ə] and [dɔ^ə]. Rule (6) also affects the glide of *there*, *their* and y*our*, a process which eventually led to phonetic forms which are homonymous with *the*y and y*ou*. Here we are concerned with the effect of (2) and (6) upon *are*:

(95)	##ar##		
	##aə##	vocalization of *r*	(2)
	##ăə##	weak word rule	(4)
	##əə##	vowel reduction	(5)
	##ə ##	loss of post-vocalic *ə*	(6)
	## ##	contraction	(9)

Contraction of *are* is therefore equivalent to deletion; there is nothing left for rule (13) to apply to; or if contraction does not apply to some forms, the deletion process will certainly eliminate them. In any case, the net result is that far fewer *are* forms survive in NNE than *is*: for many speakers, deletion of *are* is (semi-) categorical.[43] The forms of the contraction and deletion rules given here are only for *is*: in contraction (9), the lone consonant $C^1_\circ$ will show an invariance feature [*nasal] to indicate that contraction is practically total for *am*, and perhaps an additional α variable to indicate that the rule is strongly favored when the segment is $C_\circ$ rather than C^1.[44]

Rules (3) and (7) operate upon the non-central liquid *l* in a parallel fashion, so that when the auxiliary in *I will be here* is contracted, it is to all intents and purposes eliminated. There is no general process which removes the *w* in NNE or SE: a special lexical alternation is required to produce the equivalent of rule (8). This regular rule removes *h* whenever it occurs before a schwa and one or no consonant: thus the *h* in *his*, *her*, *him* is deleted as well as *have*, *has* and *had*. The form *has* is not characteristic of NNE; although there is person-number agreement in the forms of *be*, we find that the forms *do*, *have* and *was* predominate in all persons over *does*, *has* and *were*, which are not characteristic of NNE.

Contraction does not of course operate upon the pronouns his, her and him, since they do not contain the tense marker. The apostrophe used in literary conventions indicates merely the deletion of the h. Contraction does operate upon have when it contains the tense marker; in the rule given here, only the undifferentiated C_0^1 is shown for the consonant remaining. The full form will specify, as noted above, that contraction is categorical when the C contains the feature [+nasal], but that it is variable before oral consonants. The resultant ##v## will be deleted by rule (13):

(96)	##hæv##		
	##hæ̆v##	weak word rule	(4)
	## əv##	vowel reduction	(5)
	## v##	contraction	(9)
	## ##	deletion	(13)

The deletion rule (13) now shows that a lone oral continuant is removed: that is [v] and [z], but not [d] or [m]. We do not have complete data on any of the other verb forms as yet, but there seems to be little question but that the grave member [v] favors deletion more than [z]. This is particularly true, of course, before labial consonants, so that I've been would be among the rarest of NNE forms.

A word of caution is in order before accepting all of these rules as productive processes in the NNE grammars of any given speech community. In general, phonological processes are reversible: if an auxiliary disappears through the vocalization of /r/, it can re-appear if that phonological rule no longer operates or is reversed. But it appears that irreversible change can take place when phonological change identifies one lexical item with another so that the underlying forms alter. This may indeed be the case with *they book* or even with the zero form of *We crazy*, for some speakers. In the first case, we find that rules (2) and (6) operate upon the underlying possessive as follows:

(97) ðe+r
ðe+ə vocalization of *r* (2)
ðe+ loss of post-vocalic *ə* (6)

The last item falls together with [ðe˅^{I}], the phonetic output of the pronoun *they*, and even when rule (2) is strongly restricted, the form [ðe˅^{I}] may be used in attributive position: in effect, speakers may have re-analyzed the phonetic form as equivalent to that which appears in subject position as *they*. Despite the fact that the absence of a possessive /z/ suffix may reinforce this analysis, it is clear we are dealing with what was, originally at least, a phonological process: in Southern white dialects which use dummy *there* in *There's a difference*, the form /ðer/ undergoes the same process to produce a phonetic form equivalent to *they*, without any involvement of the possessive category. The extent to which such lexicalization has taken place is a topic ideally suited for empirical study through the techniques of accountable, quantitative investigation outlined above.

These brief notes on verb forms other than *is* are not intended to give a definitive account of their treatment in NNE; that is not possible without the same type of quantitative data which we have supplied for *is*. This broader view of the operation of the system allows us to show how the rules for contraction and deletion of *is* are embedded in a more general set of processes which govern the phonetic form of the NNE verbal system. The construction of such broader rules raises questions which can be resolved by more detailed investigations of variable rules. For example, closer study of the relation of v-deletion to z-deletion will allow us to determine whether the comparative infrequency of the *have* perfect in NNE (as compared to the relatively common *had* pluperfect) is due to phonological processes or to less frequent use of the grammatical category itself.[45] Preliminary investigations of *are* indicate that the low frequency of full or contracted forms is coupled with a complementary excess of the invariant form of *be* in the same contexts: *they be with us all the time*, *you be foolin' around*--far more frequent than in contexts where *is* normally appears. If this quantitative relation is as regular and reproducible as those considered above, we will be forced to modify our view of the 'habitual' or 'iterative' semantic load of invariant *be*.[46]

In general, we find a productive pattern of research in the alternation of logical arguments in the generative tradition with empirical investigations of variable rules which can resolve decisively the questions raised.

8. *The general implications of the study of variable rules*. This paper has presented a systematic exploration of a particular problem in the grammar of NNE, using controlled data from the speech community in a formal rule system adequate for the purpose.

More generally, the paper is directed at the methodological problem which seems to me of overriding importance in linguistics at the moment: to connect theoretical questions with a large body of intersubjective evidence which can provide decisive answers to those questions. In the first statements of generative grammar, it was proposed that theories could rest upon a great number of clear cases where intuitive judgments on well-formedness were uniform throughout the community--and that the theory would then decide the marginal cases. But the number of papers based upon idiosyncratic and uncertain judgments has multiplied rapidly as the questions become sharper and the analysis more detailed.[47] No matter what help the theorist's intuitions may give him in formulating his hypotheses, it is clear that his own intuitions are the only kinds of data which are not allowable as evidence, for no one can estimate the degree to which such judgments are influenced by the universal and understandable desire to prove oneself right. In any case, the construction of complete grammars for 'idiolects', even one's own, is a fruitless and unrewarding task; we now know enough about language in its social context to realize that the grammar of the speech community is more regular and systematic than the behavior of any one individual.[48] Unless the individual speech pattern is studied within the overall system of the community, it will appear as a mosaic of unaccountable and sporadic variation.

The data that we need cannot be collected from the closet, nor from any library, public or private; fortunately for us, there is no shortage of native speakers of most languages, if we care to listen to them speak. Without such empirical data, we are now in the process of producing a great many well-formed theories with nothing to stand on--beautiful constructions with ugly feet. The test of simplicity--some internal evaluation measure which is in the continuous process of revision--has not given much satisfaction to very many linguists to date. It seems reasonable to ask that alternative analyses of the data on hand prove their value by pointing to further data which can conclusively resolve the alternatives proposed.

It seems necessary at this point to refer to the distinction between competence and performance, primarily because it is so widely discussed. I am not sure whether this is a useful distinction, in the long run. There seem to be some limitations of speakers which have to do with memory span, or difficulties in articulation, which are outside of the linguistic system proper. Surely no one would want to use the notion of performance as a waste-basket category, in which all inconvenient data on variation and change can be deposited; we have any number of labels such as "free variation", or "dialect mixture", which are readily available for this purpose. Are the variable constraints discussed in this paper limitations on performance rather than competence? For some types of consonant cluster simplification, we might be tempted to answer yes. But the variable rules themselves require at so many points the recognition of grammatical categories, of distinctions between grammatical boundaries, and are so closely interwoven with basic categorical rules, that it is hard to see what would be gained by extracting a grain of performance from this complex system. It is evident that rules (0-15) are a part of the speaker's knowledge of the language, and if some of these rules are cast in a different form than traditional categorical rules, then we must clearly revise our notions of what it means to "know" a language.

It should be equally clear that we are in no way dealing with statistical statements or approximations to some ideal or true grammar. We are dealing with a set of quantitative relations which are the form of the grammar itself. A set of rules in which all of the variable rules of (0-15) suddenly became categorical would have no direct relation to the language we have described--a number of re-organizations and striking changes in the system would be certain to take place.[49]

The study of variable rules will enable us to make progress on five general questions of linguistic theory which arise in the study of any language or speech community:

1. What is the most general form of linguistic rule? That is, what notations, conventions, schemata and interpretations allow us to account for the productive and regular patterns of linguistic behavior?

2. What relations hold between rules in a system? What principles of ordering, combination and parallelism prevail in systems such as (0-15)?

3. How are systems of rules related? What are the range of possible differences between mutually intelligible dialects; how are languages originally diverse combined within a bi-lingual speech community?

4. How do systems of rules change and evolve? This historical question is of course closely related to the last point,

5. How are rule systems acquired? How does the individual's system of rules change and develop as he acquires the norms of the speech community?[50]

This paper has been concerned with specific questions within the first and second areas, but further extensions into the third and fourth areas of investigation have been indicated at many points. The particular problem investigated here has been to determine the form and order of the rules which control the appearance of the copula and auxiliary _is_ in NNE. We began with a wide range of possible solutions: total absence of the copula; deletion of abstract _be_; deletion of the formative _is_; alternative contraction and deletion of _is_; or contraction, then deletion of a single consonant. The evidence clearly shows that the last alternative is the correct one. We combined the techniques of generative grammar with quantitative analysis of systematic variation in NNE to arrive at this result, and in so doing necessarily enlarged the conception of 'rule of grammar'. This enlargement and our methods of analysis may seem novel or even challenging to those who are convinced that linguistic theory has little to learn from the study of linguistic behavior. But we do not regard our methods or our formal treatment as radical revisions of generative grammar and phonology.

On the contrary, we believe that our findings give independent confirmation of the value of generative techniques in several ways. First, I do not know of any other approach which would allow us to work out this complex series of ordered rules, in which both grammatical and phonological constraints appear. Secondly, the stress assignment rules of Chomsky and Halle seem to yield precisely the right conditions for vowel reduction and the contraction rule. Since the contraction rule has never been presented before in detail, we must consider this independent confirmation on the basis of discrete data, clearer evidence than we can obtain from the continuous dimensions of stress or vowel reduction. We also find independent confirmation of the position and role of the tense marker, even where it takes a zero form.[51] Thirdly, we find abundant confirmation of Chomsky's general position that dialects of English are likely to differ from each other far more in their surface representation than in their underlying structures. It is possible that other theoretical frameworks can be used to present these findings with certain advantages, but this conception of ordered rules is particularly well designed to discover and display such complex sets of relations in a relatively simple way.

Cumulative and convergent results of this nature are encouraging and gratifying to us, since they confirm our general belief that inter-subjective knowledge about abstract linguistic structures is within the grasp of linguistic theory. Our aim has been more general than to solve this particular problem or to enlarge a particular theoretical framework to deal with variation. We wish to provide a model for linguistic research which will arrive at decisive solutions to theoretical questions through the use of data from the speech community. We believe that this mode of work can provide the stability and sound empirical base which is a matter of some urgency in linguistics today, and the analysis of contraction and deletion in NNE is submitted with this end in view.

FOOTNOTES

[1] The research program from which this study is drawn has been supported by the Cooperative Research Branch of the Office of Education, as Cooperative Research Projects 3091 and 3288. More complete reports of this work are available in Labov, Cohen and Robins, "A Preliminary Study of the Structure of English used by Negro and Puerto Rican Speakers in New York City", Cooperative Research Report No. 3091, Washington, D.C. 1965, and Labov, Cohen and Robins, "A Study of the Non-standard English of Negro and Puerto Rican Speakers in New York City", Cooperative Research Project No. 3288. Most of the data presented here is the result of field work by Clarence Robins and John Lewis, whose contributions to the entire study were of inestimable value. Paul Cohen of Columbia University was responsible for the largest part of the transcription of this data; the assistance of Benji Wald is also gratefully acknowledged. An abbreviated form of this paper was given at the December, 1967 meeting of the Linguistic Society of America, Chicago.

[2] More complete description of the field work and sampling procedures is provided in the final reports on Cooperative Research Projects 3091 and 3288. Exploratory work in Philadelphia, Chicago, Detroit, Cleveland and Los Angeles shows that the structure of NNE described here is essentially that of adolescent and pre-adolescent Negro youth in other northern ghetto areas; the few differences to be noted in these various areas are primarily shifts in the vowel patterns and in the use of final and pre-consonantal /r/, reflecting the regional character of the surrounding white community to some extent.

[3] The three items in brackets identify the speaker by age, peer group membership or geographical background, and tape number.

[4] See W. Stewart, "Social Dialect", Research Planning Conference on Language Development in Disadvantaged Children (New York: Yeshiva University, 1966).

[5] Examples (13-16) are drawn from the French and English Creole spoken on Trinidad as cited by Dennis Solomon.

[6] From Beryl Bailey, Jamaican Creole Syntax (Cambridge: 1966).

[7] From data gathered by Lois Bloom, Columbia University, in a longitudinal study of the acquisition of language by children 18-to-36 months old.

[8] The theoretical question involved here has been put most sharply by Chomsky, who suggests that dialects of the same language are likely to be more different in their surface structure;

and superficial aspects generally, than in their underlying representations. Those who see in NNE the influence of an underlying Creole grammar are apt to take exactly the opposite position: that certain superficial differences are symptoms of radical differences in phrase structure and organization of the grammatical and semantic categories. The general question is argued in the papers of Chomsky, Peter Rosenbaum and Beryl Bailey in *Project Literacy Report No. 2* (Ithaca, N.Y.: Cornell University, 1964). In general, our own results show that Chomsky's position is borne out in case after case; the differences between NNE and SE may be seen to depend upon differences in selections of redundant formatives in low-level segmentation transformations as in *or either*; upon subtle differences in the constraints upon particular rules, as in negative concord; and in generalizations of low-level phonological rules, as in the case to be discussed here. There are two fairly important lexical items in NNE which verge upon the status of additional grammatical categories: the habitual/iterative *be*, and the intensive/perfective *done*, but the great number of features peculiar to NNE do not reflect such differences in semantic interpretation. However, the situation may have been quite different in 18th or 19th century America, or even today with speakers heavily influenced by Caribbean patterns, as in Florida; in this respect, see some of the evidence cited by W. Stewart, "Continuity and Change in American Negro Dialects, *Florida FL Reporter*, Spring 1968.

[9] Cambridge, Mass.: 1965, p. 107.

[10] "*Have* and *be* in English syntax", *Language* 43: 462-485, 1967.

Yorktown Heights: IBM, 1967

One could also argue that there is "dialect mixture", and that one of these forms is NNE, the other an importation from somewhere--but which and where is not clear. Ain't is too deeply embedded in NNE to be so lightly removed. In the alternation of ain't and not we see a typical case of an example of the inherent variation to be discussed below; similarly, ain't and didn't alternate for the preterit in such a way that in the most excited and spontaneous vernacular of pre-adolescents, both appear in roughly equal proportions.

An interesting argument can be developed to defend the position that there is no relation between this be and a finite be that might have occurred in You good. NNE has an invariant verb be with the meaning of 'habitual', 'general' or 'iterated' action--as in I be with the Jets--you know--a lot. This verb has no alternate forms in is, am, are, was or were--it is always be, does not combine with not, and does not show any auxiliary-like properties. One could argue that the non-finite be always represents this habitual be, and that there are no modals or embedded sentences corresponding to the finite You good. Although this argument is hardly persuasive, there are many interesting issues concerning this habitual be which are beyond the scope of this paper; one is touched on in Section 8.

(44) is from a toast, a long rhymed epic of NNE oral folklore, which represents the most formal aspect of the vernacular; (45) and (46) are from interchanges in group sessions.

A summary of the possible arguments to show that there is no relation between sentences of the form (1-12) and the types (29-46) and therefore no underlying is or are in NNE, might take the following form: (a) was and were are past tense markers; (b) ain't is merely a negative marker; (c) I'm is an allomorph of I; (d) i's, tha's and wha's are allomorphs of it, that and what; (e) be is related to habitual be and not to the finite copula, (f) imperative be the same; (g) emphatic forms are imported from SE; (h) the same with yes-no questions with is and are; (i) tag questions are examples of automatic is support, parallel to do support; (j) the same for elliptical responses, comparitive ellipsis, and (?) after WH attraction. I am indebted to William Stewart for raising some of these issues in the reference cited and in personal discussion. It is true that these arguments have a certain miscellaneous character, and there is hardly any explanatory force provided for the eccentric distribution of the various forms. But it might be argued that the existence of explanations based on the deletion of is and are are only valid from an SE point of view, reflecting the fact that the Creole grammar did adopt certain forms from standard English but not others, and that there is no productive rule for NP + be + Pred in NNE. For those who do not wish to accept arguments based upon simplicity, it is always possible to argue that the language has the miscellaneous character of (a-j), as a result of certain historical processes.

It will require further data, to be submitted in the following sections, to show that these arguments do not apply to the present-day NNE vernacular being studied here.

[16] It is of course awkward to refer to "the copula" and also include the auxiliary, unless we make a decision to treat the auxiliary as a main verb of a matrix sentence. In the discussion to follow, it will appear that the same phonological processes affect both equally, yet in the final analysis the distinction between copula and auxiliary will re-emerge as a constraint which favors deletion, in the environment ___Vb.

[17] I am citing these rules in the form used by Chomsky in his 1966 lectures at the Linguistic Institute in Los Angeles, since the Sound Patterns of English has not yet appeared at this writing. Certain modifications of the basic rules seem to be required by this data, such as the weak word rule, but on the whole the contraction process provides independent and striking confirmation of the validity of Chomsky and Halle's stress assignment rules.

[18] The nuclear stress rule, as formulated, applies to any final lexical item; as Halliday has noted, this lexical item must be a member of an open class. The adverbials today and tomorrow do not receive the nuclear primary stress in the unmarked form or non-contrastive form of Tom is wild today; in this case the primary stress is on wild. Although there is general agreement on the general outlines of the nuclear stress rule, the exact specification of the right hand bracket is a difficult matter; some kind of an "X" variable intervenes between the item to receive the primary stress and the bracket, and it is clear that it is not an easy matter to specify such variables. In any case, these difficulties do not affect the main argument presented here; the adverbial now is plainly not one of the items to receive primary stress in Tom is wild now without special contrastive emphasis; and after ellipsis of wild, we do not have Tom's now parallel to Tom's wild.

[19] The term "type marker", as used in Rosenbaum's Grammar II cited above, is the initial element in the phrase structure string which carries such features as interrogative, imperative and the various modal features, as well as tense. An early transformation incorporates the type marker with the first member of the auxiliary.

[20] Although NNE is a relatively constant set of grammatical and phonological rules throughout the age range of the Thunderbirds, Cobras and Jets, there are a number of subtle changes in the structures of the rules which take place in the shift from pre-adolescence to adolescence--principally a gain in the knowledge of the underlying forms of certain words, and a cleaning up of certain phonological rules; as we will see below, some of the basic phonological constraints upon the rules do not appear in the youngest speakers. In late adolescence, there are other changes which reflect an enlargement of stylistic range, and a growing knowledge of the norms of social evaluation of speech in the community.

[21] In the case of the adults, the lower diagram shows "casual speech" as isolated in the single interviews. The criteria for determining the shift to casual style are contrastive changes in "channel cues"--pitch, volume, tempo, and rate of breathing (which includes laughter); for Negro speakers, increases in pitch range are taken as the primary criteria, relatively much more important than with white speakers.

[22] In the quantitative studies shown here, the amount of data presented varies; in these initial variables, the patterns for six different groups in two styles are shown, so that the full regularity of the variable relations may appear. In later variables, only limited portions of the available data are presented, and when certain cross-correlations are necessary, some of the categories shown here as separate are combined. Not all of the speakers in most groups have been studied completely, and there are more data available which have not yet been transcribed; though it is possible that some of this data may later lead to changes on points of our analysis, in almost every case the regular relations are so apparent that if half or a quarter of the data presented here is taken, the relationships remain constant.

[23] And third, of course, there are extra-linguistic factors such as age, sex, ethnic group, social class, and contextual style, but we will not be considering these here. Our focus is upon the relatively constant grammars of Negro boys 10-to-17 years old who are integral members of the peer groups in which the vernacular culture is maintained.

[24] If these rules are compared to algebraic expressions, we can consider that in a linear expression $y=ax + b$, the selection of the constant b represents the variable input, and the factor a the slope which relates the dependent variable y to some other variable x. Here, however, we will not have a continuous function y, but a specific series of environmental constraints which give us a characteristic profile for the application of the rule for any given individual, group or speech community. It is an extraordinary result that these profiles are essentially the same for all the peer groups studied--that is, the rule is a part of a single grammar which we can construct for this speech community.

[25] In Fig. 5, the Jets differ from the T-birds and the Cobras in the relationship between the following noun phrase and following adjectives and locatives when a noun phrase precedes, but that the relationship is the same after a pronoun. In general, we find that the differentiation between following noun phrases, on the one hand, and adjectives and locatives on the other, is not as constant from group to group as other features, although in a given group this profile does allow us to examine the specific relations between deletion and contraction. In all cases, the D/C+D line follows the pattern of the C+D/F+C+D line: instead of remaining constant as in Case 4, it rises, as one would expect in Case 1.In

the final version of the rules given in section 7, we will not attempt any single statement about the effect of the following noun phrase.

[26] For a discussion of the "categorical view" which lies behind this concept of rule, see W. Labov, "The Linguistic Variable as a Structural Unit", Washington Linguistics Review 3: 4-22, 1966 (ERIC ED 010 871). Although we are discussing the form of re-write rules in this paper, the same considerations apply to any of the formal treatments now in use, since they are based upon the conception of invariant relations between discrete, invariant, essentially and conjunctively defined categories.

[27] It is true that a variable rule cannot be checked by any one instance, and therefore it would seem to have deprived us of that principle of accountability which is the mainstay of generative grammar. The disproof of a variable rule requires the analysis of a group of utterances, for each of a small group of speakers. Fortunately, the regularity of linguistic behavior is so great that these groups can be quite small. The patterns shown here emerge reliably in sets of utterances as small as five or ten, and since they hold for almost every speaker, a group of five speakers is more than sufficient. (The preliminary data presented in The Social Stratification of English in New York City (Washington, D.C.: 1966, pp. 113-131) showed a comparable regularity). There is no doubt that the variable rules presented here show a great advance in accountability over the label "free variation". Furthermore, they depend upon a much more general and important principle of accountability which is required in the analysis of linguistic behavior: that any variable form (a member of a set of alternative ways of "saying the same thing") should be reported with the proportion of cases in which the form did occur in the relevant environment, as compared to the total number of cases in which it might have occurred. Unless this principle is followed, it is possible to prove any theoretical pre-conception by citing isolated instances of what individuals have been heard saying. Speech is perceived categorically, and linguists who are searching for an invariant, homogeneous dialect will perceive even more categorically than most. In the study of non-standard dialects, the problem is most severe. Unwanted variants will first be set aside as examples of "dialect mixture", and only the forms most different from the standard will be reported. Gradually even the linguist perceives only the marked or exceptional form, so that it is possible to report that children say I is here or I'm is here "all the time", when in fact these forms may occur with vanishingly small frequency. The principle of accountability outlined above is motivated by the conviction that the aim of linguistic analysis is to describe the regular patterns of the speech community, rather than the eccentricities of any given individual.

[28] A critical example of such ordering occurs in the rule for -*t,d* deletion (shown in rough outline in section 7). There are two major constraints which hold back the rule, usually ordered in adolescence as $(-\alpha)$ the effect of a following vowel, and $(-\beta)$ the effect of a preceding morpheme boundary (that is, clusters formed by the past tense -*ed*). In late adolescence and adult speakers, this order is often reversed, and this reversal is connected with a greater ability to decipher the meaning of the -*ed* suffix in print. See W. Labov, "Consonant Cluster Simplification and the Reading of the -*ed* Suffix", to appear in H. Levin (ed.) *Basic Studies on Reading*.

[29] We note that such ordered series have been observed in quantitative work on vowel length. House's study of vowel duration in English (*JASA* 33:1174-1178, 1961) shows a tree with voicing as the α constraint, and tenseness as β. The third variable constraint, vowel height, shows some small departures from geometric ordering, and the fourth constraint, stop vs. fricative, is not well ordered at all in relation to the others.

[30] See section 6 below for an example of the development of a phonological constraint. Shifts of variable constraints are a plausible mechanism to account for the type of linguistic change in progress shown by Gauchat in Charmey (*L'unité phonétique dans le patois d'une commune*, Halle, 1903) and by Labov on Martha's Vineyard ("The Social Motivation of a Sound Change", *Word* 19:273-309, 1963).

[31] *That*, *what*, *it*, *lot* and *one* are the chief exceptions, but the first three obey special rules discussed below to yield *is*, *tha's* and *wha's*. *One* and its derivatives are the only pronouns which would allow us to examine the deletion rule left in this class. Impersonal *one* does not occur in colloquial speech, and the other forms are not common enough to yield reliable data at this time.

[32] The "vowels" we are speaking of here are vowels in the underlying representation. At a lower level of phonetic output, they are usually represented as ending in glides or semi-vowels.

[33] The set of rules developed below show that after contraction of *is*, the resulting [z] behaves very much like the plural /Z/ in NNE, and the third singular, possessive and adverbial /Z/ of SE. An epenthesis rule will apply across inflectional boundaries and across the word boundary which separates the contracted [z] from the preceding material. Although it is possible to show the various inflectional morphemes with underlying forms of /ez/ or /es/, the parallels shown in section 7 below make the /z/ representation more reasonable and economical.

[34] This exception, the Cobras, is based upon a relatively small number of cases, and it is possible that further data will alter the picture; but in any case, voicing is not a major effect.

[35] As noted at several points in this discussion, this absence of phonological conditioning in the younger group is characteristic of the general tendency for rules to develop in this direction with age.

[36] There are individual speakers of NNE who extend the usual rules of consonant cluster simplification to extremes, and also carry further the weak tendency to delete final single consonants, thus arriving at a high proportion of CV syllables.

[37] We have thus arrived at the point farthest removed from the original suggestion that NNE has no underlying _be_ and corresponding _is_; and even the suggestion that the morpheme _is_ is deleted cannot be considered consistent with the data provided here.

[38] The position of _gonna_ is not quite as regular as that of the other constraints; in some cases, it seems as if it is a categorical feature, yet in others we find it behaving as a variable increment to __Vb. The reason seems to be that _gonna_ can be interpreted as a quasi-modal, comparable to _wanna_ and _hafta_. This is one of the many processes of lexicalization, referred to below, which intercept phonological processes and re-interpret their results.

[39] The literary convention of writing _i's_ with the apostrophe before the _s_ indicates that the unreflecting approach to this form does see this _s_ as the descendant of an original _is_. As we will see below, this is true only in the sense that the _s_ reflects the presence of the copula, but in a non-linear fashion.

[40] One might think that such schwas would be indistinguishable from reduced forms of _are_; but in NNE the amount of person-number disagreement of _is_ and _are_ is very low, and there is practically no vestige of _are_ occurring in singular contexts.

[41] That is, there are no pronouns ending in consonant plus _t_. However, if the assibilation rule is actually broader than it seems in this formulation, it may then be ordered with respect to the cluster rules presented here.

[42] The formal treatment of variable rules developed here will allow us to make much more precise statements about the relationships between dialects or systems than have previously been possible. This is not the topic of this paper, but it is worth noting here that these relations frequently show a progressive shift of variability, so that where one system has constant rules the other is variable. For example, WNS negative concord rule is variable in all speakers, but the NNE rule is constant for those who participate fully in the vernacular culture: the negative is always transferred to all indefinites within the clause. In the case of the vocalization of _r_, considered here, a more complex relationship holds. The New York City vernacular has variable _r_

in final and pre-consonantal position, and also shows variable application of the rule corresponding to (2) before a word boundary followed by a vowel, as in four o'clock, but at at a lower frequency. Rule (2) for NNE, on the other hand, applies categorically in final and pre-consonantal position, and at a very high frequency in the type four o'clock. Furthermore, it applies at a low frequency to intervocalic r within a morpheme, so that Cal and Carol, pass and Paris, can be homonyms. The rule never applies in this position for SE or WNS in New York City.

[43] The expression (semi-) categorical indicates situations in which the rule applies with very high frequency and the small percentage of the comparatively rare cases where it does not apply can hardly be considered part of the linguistic pattern; in any given case, there is no expectation that the rule will not apply. Such situations, marked by * in the rule, often mark the remains of a rule which was once productive, and are associated with change or development with age: they are therefore not without significance in the analysis of the origins or changes taking place within the dialect.

[44] One indication that this analysis of are contraction is correct is found in the fact that working-class white Southerners do omit are in such expressions as You gittin' the salad, and Cucumbers? We out of them (from the writer's own observations in Georgia and North Carolina). On the other hand, there is no evidence for white Southerners deleting is, and the intuitive responses of a number of Southern linguists and laymen are that this is not possible for a white speaker. This is not an arbitrary selection of are rather than is, but rather a reflection of the fact that white Southerners do occasionally use rule (6) to yield po', etc.(seemingly in the same stylistic contexts as the absence of are), but have no deletion rule for is.

[45] Past perfect auxiliary had is quite common among NNE speakers, even very young pre-adolescents, especially in narrative. Have is not as frequent; and some writers have even suggested that there is no have + en in NNE. However, it will turn up readily in the environments outlined for is and are in (29-46) in section 1. For example, on the first few feet of a tape of pre-adolescent boys, recorded in Washington at the Center for Applied Linguistics, we hear "Why have you borrow my pencil?"

[46] That is, be is showing evidence of a hole-filling function. Where phonological processes eliminate one form, NNE speakers seem to be dropping in another form which is immune to those processes.

[47] Among Chomsky's first published statements on this point, we read "In many intermediate cases we shall be prepared to let the grammar itself decide, when the grammar is set up in the simplest way so that it includes the clear sentences and excludes the clear non-sentences" (Syntactic Structures,The Hague, 1955, p.14). It should then be possible to avoid presenting intermediate cases as evidence. However, a great many recent arguments in syntax have hinged upon sentence types which are evidently intermediate in grammaticality, in the sense that there is widespread disagreement or "variation" in judgments on grammaticality. One such case was critically involved with

the argument for employing a category *by + passive* as a manner adverbial, thus supporting the general argument that all transformations be obligatory and preserve meaning. The same middle verbs which do not permit the passive transformation are also said to be incompatible, or at least not combine freely, with manner adverbials. Yet we have typically (?) *the suit fitted me with a bang/splendidly/in a curious manner*; (?) *John married Mary with a bang/splendidly/in a curious manner*. It can hardly be said that the theory is here used to decide these intermediate cases, but rather that they are being used to decide the theory. (*Aspects of the Theory of Syntax*, Cambridge, 1967, p. 103). More recently, some writers on transformational grammar have asserted vigorously that although no one can be expected to agree with their judgments on grammaticality, they are describing these judgments and nothing else. It is most unsatisfactory for the author to argue that he is describing one particular idiolect, because the reader is then deprived of any possible way of evaluating the evidence; his own agreement or disagreement with all or any of the examples thus becomes irrelevant, and he becomes a passive spectator of a description which can never be validated. The linguist clearly intends to describe the structure of English, or of a particular dialect of English: if there is a speech community where the rule in question is not intermediate, but one of the clear cases, it seems reasonable to ask the investigator to establish this. Otherwise, the techniques employed here to deal with inherent variation may prove applicable to these cases of marginal grammaticality, where speakers' judgments vary according to some unknown constraint.

[48] This conclusion is documented in some detail in W. Labov, *The Social Stratification of English in New York City* (Washington: CAL, 1966).

[49] One example of such a re-organization can be seen in modern Scots, where the simplification of *-t,d* clusters after stops is categorical. In most dialects, the preterit is preserved by a re-ordering and re-structuring of the epenthesis rule, so that after stops we have *frichtit* [frɪxtət] 'frightened', *gairdit* [gerdət] 'guarded', etc. (W. Grant and J. Dixon, *Manual of Modern Scots*, Cambridge 1921).

[50] The theoretical problems outlined here are not at all irrelevant to some immediate problems of applied linguistics in teaching the reading and writing of SE to speakers of NNE. Although the primary obstacles in the schools are social and cultural factors, there are some linguistic differences which have profound effects--not because NNE is so different from SE, but because it is so similar. The conclusion reached in this

paper should make it immediately evident that the task is not so much to inhibit deletion as to teach contraction to NNE speakers--not the abstract contraction rule, but rather the control of contraction without immediately ensuing deletion. There is no English program currently in use which focuses on this critical point, since it would never occur to an SE or WNS speaker that contraction needs to be taught. When an NNE speaker says "He wild", the teacher would normally correct with "He *is* wild", thinking that this is the equivalent translation. But as we have seen, the NNE speaker would have said *He is wild* if that is what he meant. What he intended to say is equivalent to SE *He's wild*, and that equivalence must be explicitly taught. When it comes to reading, NNE speakers have a great deal of trouble with printed contractions. In the commendable desire to make primers less formal, some authors have begun to insert contractions *I'll*, *We're*, without realizing what difficulties they are creating for NNE readers, for whom full forms *I will* and *We are* are perfectly natural--much more so than for WNS readers. Thus in more than one way a knowledge of the abstract rule system of NNE is essential for the right approach to educational problems.

[51] There are a great many other ways in which this data feeds back into and is informed by general problems of syntactic theory. For example, there is one apparent exception to the rules which relate contraction to deletion as seen in sentences of the type:

What I mean by bein' destroyed, they was brought up into they rightful nature.	[29, N.J., #737]
All I knowed, that I was in the hospital.	[13, T-Birds, #458]
All I could do, as' him what he's tryin' to do.	[16, NYC, YH 33]
But next thing I knew, he was on the ground.	[16, Jets, #560]

These sentences are anomalous in two ways: (1) SE does not contrast in this position--we do not have, *What I mean's you're crazy*, and (2) some white speakers do delete, and can produce sentences similar to those above. The fact that these sentences are exceptions in *both* ways clearly indicates that they are not true counter-examples to the generalizations of section 3. The problem lies in the analysis of the derivation of these sentences: to this point, there has been little discussion and no agreement on the underlying structure of cleft sentences such as *What I mean is you're crazy*. There are clearly two major possibilities: (a) derivation from [WH-indef [I mean WH-indef]$_S$ - COP - S]$_S$, in which the *is* is the main verb of the sentence, and (b) derivation from [I mean WH-indef [WH-indef - COP - S]$_S$]$_S$ where *mean* is the main verb of the sentence, and the *is* is the verb of a relative. It is the second analysis which seems to be followed by speakers who delete this *is*, using the same deletion transformation involved in *The book that is yellow with age* → *The book yellow with age*... which has nothing to do with the contraction and deletion rules discussed here. If this is the case, we can understand how both NNE and white speakers can delete this item, and that its uncontractability stems from its immunity to vowel reduction at later stages of the grammar. This particular possibility seems to involve primarily verbs of knowing, saying and meaning and also the pro-verb *do*; we do not get *What I broke, my leg*, and there is much more to be said about the implications of this fact. A fuller discussion is given in CRP 3288.

[17a] Since The Sound Pattern of English has now appeared (New York: Harper and Row, 1968) it is possible to contrast Chomsky and Halle's treatment of "weak words" with that given here. There is no explicit discussion of the copula, but it appears that primary stress will not be assigned to auxiliaries by the normal process; in this respect, the auxiliaries are treated like articles, conjunctions and prepositions. The main stress rule applies before brackets labelled with major categories: N, P. S, V or A (p. 240). The # boundary is automatically inserted at the beginning and end of every string dominated by a major category (p. 366): thus we have surface structures such as $[_S\#[_N\#\mathrm{John}\#]_N[_{VP}\#\mathrm{is}[_A\#\mathrm{wrong}\#]_A\#]_{VP}\#]_S$. The copula is thus not a "phonological word" in the technical sense, and does not receive main stress; the vowel of *is* is automatically reduced unless contrastive stress intervenes. (It is likely that the copula and auxiliaries, as well as prepositions, would not appear in the deep structure, but be spelled out transformationally from features of lexical items or type markers, as in Rosenbaum's *Grammar II* [Yorktown Heights: IBM, 1968].) This treatment provides a simpler mechanism than the "weak word rule" given above. After the deletion transformation, we have $[_S\#[_N\#\mathrm{John}\#]_N[_{VP}\#\mathrm{is}\#]_{VP}\#]_S$; the same rules which give primary stress to *wrong* without deletion, give primary stress to *is* after deletion. We are thus able to do away with the special feature [+W] and predict the behavior of *is* and *was* from very general rules.

Test cases for the operation of this mechanism would be provided by forms where adverbs follow the auxiliary in a truncated sentence. In *John is never wrong* the adverb is not in construction with the auxiliary, but rather with the verb phrase as a whole. The mechanism we have proposed would predict that after deleting *wrong* the copula *is* would remain unreduced, while the general stress rule of Chomsky and Halle would stress the last element of the remaining verb phrase and leave the first one without stress. The difficulty here is that such truncated phrases regularly switch the adverb and the copula, so that *John is never wrong* → *John is never* → *John never is*. There is some tolerance for the unreversed form *He is always*; here the stress pattern (if acceptable) is that predicted by the weak word rule; **He's always* is definitely not possible. But the clearest evidence is derived from strings of auxiliaries in truncated sentences: *He has been*; *He may have*; *He may have been*; *He will have been*. The Chomsky-Halle stress rule as it now stands would assign main stress to the last item of such surface structures, yielding *$\overset{2}{\text{He's}}\ \overset{1}{\text{been}}$; *$\overset{2}{\text{He}}\ \overset{-}{\text{may}}\ \overset{1}{\text{have}}$; *$\overset{2}{\text{He}}\ \overset{-}{\text{may}}\ \overset{-}{\text{have}}\ \overset{1}{\text{been}}$; *$\overset{2}{\text{He'll}}\ \overset{-}{\text{have}}\ \overset{1}{\text{been}}$. But it is clear that primary stress is assigned instead to the first auxiliary in the string in the unmarked case. Furthermore, the other items are not fully reduced: we have secondary or tertiary stress in such patterns as $\overset{2}{\text{He}}\ \overset{1}{\text{has}}\ \overset{3}{\text{been}}$; $\overset{2}{\text{He}}\ \overset{1}{\text{will}}\ \overset{3}{\text{have}}\ \overset{3}{\text{been}}$. It seems clear that stress has been assigned to all members of the auxiliary at some point in the derivation, and that the cyclic operation of the main stress rule does not reduce them to the point that allows vowel reduction and contraction. There must be some provision in the stress rules for supplying stress to such auxiliaries--either by the main stress rule or some supplementary rule. Furthermore, the rule we have suggested for eliminating stress in "weak words" will be part of a set of adjustments for reducing vowels which have at one point received stress. Chomsky and Halle state that the second vowel in *condensation*, which receives stress in the first cycle, is never reduced, as opposed to the second vowel of *compensation*, which is reduced. There is clearly a difference in the treatment of these two vowels in careful speech, but it is also true that the second vowel of *condensation* reduces to schwa in the more casual treatment of many speakers. Finally, we should note that the category of "weak words", with the feature [+W], is independently motivated by the rules for tensing and raising of short *a* in many dialects of English. As Paul Cohen has pointed out, it is this feature which allows us to generate in a single rule the oppositions *an* ~ *Ann*, *have* ~ *halve*, *can* [Aux] ~ *can* [Verb, Noun], *as* ~ *razz* etc., which are so common in Eastern dialects from New York to Philadelphia and elsewhere.

[27a] We may also need conventions which will indicate that the presence of a given feature prevents the rule from applying, i.e., that $\varphi = 0$. Present conventions only permit us to insert $[-\text{fea}_i]$ for this purpose as an environment governing the rule. But the positive notation is needed when such a condition develops in the course of linguistic evolution as the limiting case of a trend in which the presence of $[\text{fea}_i]$ interferes with, rather than promotes, the operation of the rule when we are as close to zero as the semi-categorical rule is close to one. Such a notation Δ will be interpreted as $\varphi = \left(\frac{-1\Delta 1}{-2}\right) P$ where P is the expression for the evaluation of φ given as (76) above.